

Modeling CLV: A test of competing models in the insurance industry

Bas Donkers · Peter C. Verhoef · Martijn G. de Jong

Received: 30 August 2005 / Accepted: 28 November 2006 /
Published online: 5 June 2007
© Springer Science + Business Media, LLC 2007

Abstract Customer Lifetime Value (CLV) is one of the key metrics in marketing and is considered an important segmentation base. This paper studies the capabilities of a range of models to predict CLV in the insurance industry. The simplest models can be constructed at the customer relationship level, i.e. aggregated across all services. The more complex models focus on the individual services, paying explicit attention to cross buying, but also retention. The models build on a plethora of approaches used in the existing literature and include a status quo model, a Tobit II model, univariate and multivariate choice models, and duration models. For all models, CLV for each customer is computed for a four-year time horizon. We find that the simple models perform well. The more complex models are expected to better capture the richness of relationship development. Surprisingly, this does not lead to substantially better CLV predictions.

Keywords Customer lifetime value · CLV-models · Forecasting · Database marketing

JEL Classification M30 · C53 · C35

B. Donkers (✉)

Department of Business Economics, Erasmus University Rotterdam, Office H15-19,
P.O. Box 1738, NL-3000 DR Rotterdam, The Netherlands
e-mail: Donkers@few.eur.nl

P. C. Verhoef

Faculty of Economics, Department of Marketing, University of Groningen,
P.O. Box 800, NL-9700 AV Groningen, The Netherlands
e-mail: P.C.Verhoef@rug.nl

M. G. de Jong

Department of Marketing Management, RSM Erasmus University, Office T10-17,
P.O. Box 1738, NL-3000 DR Rotterdam, The Netherlands
e-mail: MJong@rsm.nl

1 Introduction

Customer lifetime value (CLV) has become an important metric in marketing and specifically in customer relationship management (Rust et al. 2000). The increase in attention for CLV fits well in the emerging literature on customer behavior and customer profitability (Hogan et al. 2002). One of the key issues when firms use the CLV-metric is whether the firm can provide an adequate prediction of the CLV of each customer in their database (Malthouse and Blattberg 2005; Venkatesan and Kumar 2004). Based on these predictions, firms can decide on their investments in (segments of) customers (Zeithaml et al. 2001). Also, a summation of CLV predictions of all customers results in a valuation of the total customer base, which can be important for firm valuation purposes (Gupta et al. 2004).

A customer's CLV is typically assessed using customer behavior data from the company's database to predict customer behavior and profitability. The traditional direct marketing literature (Berger and Nasr 1998; Jain and Singh 2002) has proposed rather simple models, mainly using aggregated data (e.g., aggregate retention rates) to predict CLV. In the general marketing literature, researchers have proposed behavior-based models (e.g. logit-models, multivariate Probit models) to predict customer behavior, including cross-buying and retention (Kamakura et al. 1991, 2003, 2004; Knott et al. 2002). Finally, there are several studies using duration models to quantify purchase incidence (Jain and Vilcassim 1991), which in combination with a purchase quantity model, enable assessment of CLV at the individual level (Fader and Hardie 2001; Schmittlein et al. 1987).

So far, there are no studies that provide an overview and an empirical comparison of a large set of models—varying in realism and complexity—to predict CLV (Kamakura et al. 2005). Only Malthouse and Blattberg (2005) provide a comparison of two different models for CLV prediction. At the same time, research on modeling CLV is one of the MSI research priorities (MSI 2004). In this study we consider the prediction of CLV in multi-service industries, in particular in the insurance industry, where purchase behavior can be especially complex. In multi-service industries, customer behavior is multi-dimensional: not only customer retention, but also cross-buying and service usage, are important components of CLV (Bolton et al. 2004). Accounting for these behaviors at the individual level results in more complex, but also more realistic models. The question is whether increasing model complexity also yields a better predictive performance. If not, managers can feel comfortable relying on rather simple models.

In our modeling approach, we consider both relationship-level and service-level models. The relationship-level models focus on customer retention and profits aggregated across services. The service-level models take a disaggregated perspective, and can also account for cross-buying of services. Within the class of service-level models, we consider models that assume independent purchase decisions, as well as models with unobserved heterogeneity that might result in dependencies across the decisions for the various services (Kamakura et al. 1991; Li et al. 2005; Manchanda et al. 1999). The dependencies are incorporated through a latent factor structure (Kamakura et al. 2003) allowing a parsimonious specification of this complex phenomenon.

Our models in fact do not pertain to CLV prediction but to the modeling and prediction of next period's customer behavior. Subsequently, CLV is derived from the predicted behavior. For the simplest models, prediction of CLV is straightforward. For the service-level models with explanatory variables, we use a Markov model to predict more than a single period ahead (Pfeifer and Carraway 2000; Rust et al. 2004). CLV predictions by means of Markov models have not received much attention in the CLV literature. Hence, we provide additional details on this approach and the associated complications.

After estimating the models and computing the corresponding CLV predictions, we compare the predictive performance of these models using a longitudinal dataset of customer behavior for an insurance company. Comparisons of the predictive performance are made in three domains: (1) predicting the level of individual CLV, (2) predicting the ordering of customers based on CLV, and (3) valuing the total customer base. These different types of predictions relate to different tasks marketing managers have to perform, including budgeting, segmenting, and firm valuation. To better understand the results, we also compare the predictive performance of these models after 1 year and after 4 years.

One of the most prominent findings of our comparison is that simple models perform relatively well. These simple models, however, need to account for both customer retention and cross-buying. The models that only account for customer retention fail to capture an important component of a customer's value and consequently predict poorly. Although these findings are industry specific, they might resonate beyond the insurance industry due to the nascent nature of CLV research.

2 Modeling CLV

The basic formula for calculating CLV for customer i at time t for a finite time horizon (T) (Berger and Nasr 1998) is

$$CLV_{i,t} = \sum_{\tau=0}^T \frac{Profit_{i,t+\tau}}{(1+d)^{\tau}}, \quad (1)$$

where d is a pre-determined discount rate. In multi-service industries, $Profit_{i,t}$ can be defined as

$$Profit_{i,t} = \sum_{j=1}^J Serv_{ij,t} * Usage_{ij,t} * Margin_{j,t} \quad (2)$$

Here J is the number of different services sold, $Serv_{ij,t}$ is a dummy indicating whether customer i purchases service j at time t , $Usage_{ij,t}$ is the amount of that service purchased, and $Margin_{j,t}$ is the average profit margin for service j . In line with the business practice of the company in our application, we use margins that are not customer specific. According to the company, obtaining accurate estimates of individual-level profit margins is not possible for the insurances they sell as this requires an accurate assessment of the individual-level risks that are insured.

In our application we consider models at two levels: (1) relationship-level models and (2) service-level models. The relationship-level models focus on relationship length and total profits, building directly on Eq. 1. The service-level models disaggregate a customer's profit into the contribution per service. CLV predictions are then obtained by predicting purchase behavior at the service level and combining Eqs. 1 and 2 to forecast CLV. When analyzing service level purchases, the dependent variable will be a zero–one variable indicating whether the service was purchased in a given year. This is in line with the annual contracts used in the Dutch insurance market. More important, these contracts cannot be terminated before the end of the contract year, so purchasing an insurance policy means that the customer pays premium for the full 12 months. In other settings, usage might also take on values below one as contracts can be ended during the year. In such instances, this obviously should be modeled appropriately.

2.1 Overview of models for customer behavior

We consider both simple and more complex customer behavior models. At least two views can be taken on the decision process underlying the purchase behavior for services. The first view assumes that customers evaluate their particular service

Table 1 Overview of rival models

Model	Mathematical model
1. Relationship-level models	
1.1. Status quo	$\text{Profit}_{i,t+1} = \text{Profit}_{i,t}$
1.2. Profit regression	$\text{Profit}_{i,t+1} = \alpha_0 + \alpha_1 \text{Profit}_{i,t}$
1.3. Retention (segmented)	
Aggregate retention rate	$E_t\{\text{Profit}_{i,t+1}\} = P_t(\text{ret}_{t+1})\text{Profit}_{i,t}$
Segmented retention rate	
Recency of last purchase/ cancellation	
Frequency (number of insurances)	$E_t\{\text{Profit}_{i,t+1}\} = P_t(\text{ret}_{t+1} \mid \text{segment}_i)\text{Profit}_{i,t}$
Monetary value (high vs. low profit)	
RFM-index (high vs. low)	
1.4. Probit model	$E_t\{\text{Profit}_{i,t+1}\} = P(\text{ret}_{t+1} \mid x_i)\text{Profit}_{i,t}$
1.5. Bagging	$E_t\{\text{Profit}_{i,t+1}\} = P(\text{ret}_{t+1} \mid x_i)\text{Profit}_{i,t}$
1.6. Duration model	$E_t\{\text{Profit}_{i,t+1}\} = P(\text{ret}_{t+1} \mid x_i, \text{Duration}_{i,t})\text{Profit}_{i,t}$
1.7. Tobit II model	$E_t\{\text{Profit}_{i,t+1}\} = \Phi(x_i' \beta)(\alpha_0 + \alpha_1 \text{Profit}_{i,t} + E\{\epsilon_{i,t+1} \mid \text{retention}\})$
2. Service-level models	
2.1 Independent choice models	$E_t\{\text{Profit}_{i,t+1}\} = \sum_{j=1}^J P(\text{purchase}_{i,t+1,j} \mid x_i) \text{Margin}_j$
2.2 Independent duration models	$E_t\{\text{Profit}_{i,t+1}\} = \sum_{j=1}^J P(\text{purchase}_{i,t+1,j} \mid x_i, \text{Duration}_{i,t}) \text{Margin}_j$
2.3 Multivariate choice models	$E_t\{\text{Profit}_{i,t+1}\} = \sum_{y \in Y} P_t\{\text{portfolio}_{i,t+1} = y \mid x_i\} * \text{Margin}_y$
2.4 Multivariate duration models	$E_t\{\text{Profit}_{i,t+1}\} = \sum_{y \in Y} P_t\{\text{portfolio}_{i,t+1} = y \mid x_i, \text{Duration}_{i,t}\} * \text{Margin}_y$

contract(s) periodically, resulting in a discrete choice model for purchases in that period. The second view stems from critical-incident thinking, which is prevalent in the service literature (Gremler 2004; Keaveney 1995). It is assumed that customers re-evaluate their service contract(s) only after some critical incident takes place, for example, the insurance company denies a claim. The corresponding model is a duration model (Fader and Hardie 2001; Reinartz and Kumar 2000), which models the time until a critical incident occurs that results in a non-renewal of service(s). Note that critical incidents typically refer to events that are under control of the company, e.g. inappropriate service deliveries. In our analysis of customer retention, we will model the arrival of relationship terminating events, which could be critical incidents in the service process but also an offer from a competitor that is accepted by a customer.

In this section we introduce the customer-behavior models, classified into relationship-level and service-level models. Table 1 provides an overview of the models and their mathematical formulation. Technical details on model specification and estimation are deferred to an appendix. The section ends with a discussion of the common set of explanatory variables we include.

2.1.1 Relationship-level models

The relationship-level models do not explicitly model buying behavior of individual services, but rather consider retention and total profits aggregated across services. We consider six types of CLV models in the class of relationship-level models: a status quo model (1.1), which assumes that profits simply remain constant over time; a regression model (1.2), which aims at predicting a customer's annual profit contribution (Malthouse and Blattberg 2005); customer retention models (1.3) using either aggregate or segment-level retention rates, where we consider segmentation along the familiar Recency, Frequency and Monetary value dimensions (Berger and Nasr 1998; Rossi et al. 1996); models for customer-specific retention probabilities, either obtained through a Probit model (1.4) (Bolton et al. 2000), or through a bagging approach (1.5) (Lemmens and Croux 2006); duration models for a customer's relationship duration (1.6) and finally a Tobit II model (1.7).

For all models that focus on customer retention or relationship duration only, the assumption is that profits, conditional on retention, remain unchanged. The expected profit in period $t+1$ can thus be defined as

$$E_t\{\text{Profit}_{i,t+1}\} = P_t(\text{Ret}_{i,t+1})\text{Profit}_{i,t} \quad (3)$$

The customer relationship duration model (1.6) requires a bit more exposition given the traditional use of duration models in non-contractual rather than contractual settings. In the CLV literature there is a large stream of research based on the Pareto/NBD framework. As Schmittlein et al. (1987) already indicated (p.16) “the Pareto/NBD is inappropriate for examining renewable service contracts,..., since the opportunities for transactions occur at regular, observed intervals.” Moreover, in our application purchase quantity is always one, rendering the Pareto/NBD framework inappropriate for our application. As said, however, we do use a duration model, inspired by the frequent application of the Pareto/NBD framework in the CLV literature.

The major difference between a contractual setting and the typical application for the Pareto/NBD framework is that the end of a relationship is observed in a contractual setting and need not be inferred indirectly from customer behavior. Depending on the type of service contracts, the contract can be cancelled either at the end of the contractual period or at any moment in time. In the latter case, the time until the end of the relationship can be modeled and estimated straightforwardly with a hazard rate model as is done in the Pareto/NBD framework. In our application we have renewable annual service contracts that cannot be ended before the end of the year. As a result, we can only observe whether an event that is detrimental to the relationship has occurred within a given year, not the exact timing of that event within that period. Details on how the estimation procedure for duration models is adjusted to account for this can be found in Appendix A, but see also Meyer (1990).

The rate at which relationship-terminating events occur—the hazard rate—is likely to depend on covariates such as types of services an individual purchases. Therefore, we use a proportional hazard model (Cox 1972; Fader et al. 2004) to include explanatory variables in the hazard rate. The major advantage of the duration model approach is that it permits modeling of the phenomenon that customers with a long relationship are more loyal (Bolton 1998). As a result, not only the explanatory variables in the proportional hazard model, but also the customer's relationship duration are central ingredients of the predicted retention probabilities.

For the shape of the hazard rate we consider two alternatives. As a flexible parametric specification for the baseline hazard we use the quadratic Box–Cox formulation proposed by Helsén and Schmittlein (1993). We also consider a fully non-parametric baseline hazard rate¹ as is proposed by Prentice and Gloeckler (1978), see also Meyer (1990).

Also the Tobit II model has not been used very often, but see Hansotia and Wang (1997) for an exception. The potential strength of this model is that profits are only considered in case of retention, separating out the effect of customer defection on profitability.

2.1.2 Service-level models

In this section, we introduce the models that concern service-level purchase behavior. Such a disaggregated approach might be preferable to relationship-level approaches. Since margins per service vary across the services and not across customers, the only variable of interest at the service level is whether the service is purchased or not. We consider two types of models: binary Probit models and duration models.² Although not obvious at first sight, there is a major difference between the choice model approach and the duration model approach to service purchase behavior. The choice model approach has as dependent variable the

¹We thank an anonymous referee for suggesting this non-parametric approach.

²As the number of observations per service is rather small for some services, we only estimate parametric hazard models.

decision to purchase a service or not. The duration models focus on the duration of an existing relationship, so the decision to continue purchasing a service is modeled.³ So, in contrast with the choice models, the duration models only model the ending of the period in which a service was purchased, not the starting of a new one.

For both choice and duration models, we consider independent and dependent service-level equations. This leads to a model with J independent Probit models (Model 2.1), a model with J independent duration models (Model 2.2), a multivariate Probit model (Model 2.3), and a multivariate duration model (Model 2.4). For dependent equations, we can incorporate unobserved heterogeneity, affecting the purchase behavior for multiple services. Simultaneous modeling of the J purchase decisions permits estimation of customer-level heterogeneity, because there are multiple observations per customer (cf. Li et al. 2005). Both firm and customer behaviors can drive the dependencies in purchase decisions across services. For instance, firms offer package deals, making it attractive for customers to buy a certain combination of services. From the customer perspective interdependencies might arise from customer satisfaction, perceived switching costs, etc.

We take a general approach to incorporate dependencies across equations building on the ideas in Kamakura et al. (2003). More specifically, the dependencies across the decisions on the J services result from an unobserved error component $\varepsilon_{i,t}$ (of dimension J) that is added to the systematic component, $\mathbf{x}'\beta$, for both the duration models and the Probit models. This unobserved component is characterized by a factor structure: $(\varepsilon_{i1,t}, \dots, \varepsilon_{iJ,t})^T = \Lambda \mathbf{f}_{i,t}$. Here, $\mathbf{f}_{i,t}$ denotes the vector of unobserved factor scores and Λ the matrix with factor loadings, which together allow for dependencies between the purchase decisions. This results in a parsimonious representation of all kinds of possible dependencies. More details are provided in Appendix A.

We only allow for unobserved heterogeneity in the service-level models, as one needs very strong and untestable assumptions to identify unobserved heterogeneity from a single observation per observational unit (customer in our case). A solution at first sight would be to use 2 years of data for model estimation instead of only one. However, this does not solve the identification problem for the models that focus on customer retention, or equivalently customer relationship length. With these models, one cannot distinguish between unobserved heterogeneity and duration dependence.⁴ Intuitively this is clear from the fact that only a single observation on the relationship duration is available; adding a single year only increases the informational content of this single observation. For the models that include the prediction of profits, using an additional year of data does result in identification of unobserved heterogeneity. This is, however, unlikely to be worthwhile as almost all variation is captured by the level of past profits and adding unobserved heterogeneity is unlikely to improve the predictive performance.

³In line with the usual definition of relationship length used in other types of market research (Bolton 1998), we use the length of the relationship with the company as the relevant duration for all services.

⁴We thank an anonymous reviewer for this insight.

2.1.3 Predictors of customer behavior

The models discussed above all aim at predicting (part of) customer behavior. In line with the traditional CLV-literature (e.g. Berger and Nasr 1998), we only include past behavioral data available from the customer database to predict customer behavior. Recency, frequency, and monetary value of a customer have proven to be powerful predictors (Rossi et al. 1996). We include dummies for ownership of each insurance type in the previous year, which represent purchase frequency, but also profit (monetary value), as profit is the sum of the purchase dummies for each type of insurance times the insurance specific margins. For recency we include two dummy variables; one dummy indicating whether the customer purchased a new service the last year (purchase recency) and one dummy indicating whether the customer cancelled a service last year (cancellation recency). We also include relationship age as predictor of customer behavior. In the regression-type models we include two dummies for the first and second year of the relationship. The duration models include relationship length only through the hazard rate. Finally, loyalty program membership is included as a predictor in the models (Bolton et al. 2000).

2.2 Predicting CLV

For the relationship-level models, it is intuitively clear how to extend profit predictions further into the future. For example, for the retention models the likelihood of a customer being retained 2 years in a row is the retention probability squared. Formally, this intuition mimics a first order Markov chain for the prediction of profitability at time $t+2$, given the predictions for $t+1$ (Pfeifer and Carraway 2000; Rust et al. 2004). To predict more than one period ahead with the relationship-level models, however, we need to assume that the explanatory variables in the relationship remain constant. Alternatively, one could model the dynamics of the explanatory variables, as is done in the service-level models.

For the service-level models, we have discussed how to predict customer behavior, but still need to establish the link between customer behavior and profits. As the marginal probabilities of purchasing an insurance follow directly from the model, predicting profits in the first year is straightforward. For subsequent years, we can use the information on the predicted portfolio of services being purchased to update the explanatory variables, potentially improving the predicted purchase probabilities for all services. In doing so, we do not only need the marginal probabilities of purchasing a service but also the probabilities of owning a given portfolio of services, which we denote by $y_{i,t} = (y_{i1,t}, \dots, y_{iJ,t})^T$. Let Y denote the set of all possible portfolios, then the corresponding expected profits of a customer are given by

$$E_t\{\text{Profit}_{i,t+1}\} = \sum_{y \in Y} P_t\{\text{portfolio}_{i,t+1} = y\} * \text{Margin portfolio } y \quad (4)$$

The insurance-level models are able to generate a complete picture of the customers' behavior over time. These models thus enable the prediction of all explanatory variables, including insurance ownership and the recency variables. In this setting with the changes in the explanatory variables explicitly modeled, a

Markov chain can also be used. This, however, requires using an extensive state space that contains all the variables in the model.

We now discuss in more detail how the extensive Markov chain for the insurance-level models can be constructed. In total, there are J services, so there exist 2^J possible portfolios of services purchased. Adding two recency dummies, a dummy for membership of the loyalty program, and relationship age truncated at age L , the total number of states equals $2^{J+3} * L$, which is the number of portfolios times the four possible combinations for the recency dummies times the two levels of the loyalty program times the number of possible relationship age levels. For other applications, other state spaces have to be constructed, but the methodology extends the work by Pfeifer and Carraway (2000) and is generally applicable. Note that we do not consider the Markov chain methodology a separate CLV model. In our view, Markov chains provide a tool to compute CLV for models that describe behavior only for one period ahead, such as the retention models and the insurance-level models. However, the Markov chain methodology plays a crucial role in CLV prediction, warranting further discussion.

We illustrate the working of the Markov chain with a simplified setting. Consider the situation with ownership of only two service types, relationship duration and the two recency dummies as the variables describing the state space. Considering only relationships of 1 period or 2 periods, the number of states is 2^2 possible ownership combinations times 2 relationship durations times 2^2 possible combinations for the recency dummies, a total of 32 possible states.

The number of states that can be reached from a certain state, however, is, in general, much smaller. For example, one cannot go from a state with relationship duration j to one with relationship duration $j - 1$ or $j + 2$. Many other restrictions are more subtle and depend, for example, on the value of the recency dummies. As a simplified example, consider the states a customer can reach in period 2, when in period 1 s/he purchases service 1 and not service 2. Note that the values of the recency dummies do not matter here, as they do not affect whether a state can be reached. Table 2 presents this current state and the 16 states that are possible in period 2, as states that have $t = 1$ are not reachable. As is clear from this table, only four states can be reached. The actions the customer has to take to reach each of these four possible states are:

- State 2: Customer starts purchasing service 2, so recency buy equals 1. He also continues buying service 1.
- State 8: The customer purchases the same services. Recency dummies all equal zero.
- State 9: The customer stopped purchasing service 1, so recency quit equals 1, and started purchasing service 2, so recency buy also equals 1.
- State 15: The customer stopped buying services from the company. Recency quit equals 1.

The state space for the purchase decisions on a larger set of services is much larger, in our application there are more than 90,000 states, but similar restrictions hold on the possible transitions from one state to another. The number of transitions that occur with positive probability is therefore large but manageable and the

Table 2 Possible transitions to future states

State	Period	Ownership service 1	Ownership service 2	Recency buy	Recency quit	Reachable Y/N
Current	1	1	0	–	–	
1	2	1	1	1	1	N
2	2	1	1	1	0	Y
3	2	1	1	0	1	N
4	2	1	1	0	0	N
5	2	1	0	1	1	N
6	2	1	0	1	0	N
7	2	1	0	0	1	N
8	2	1	0	0	0	Y
9	2	0	1	1	1	Y
10	2	0	1	1	0	N
11	2	0	1	0	1	N
12	2	0	1	0	0	N
13	2	0	0	1	1	N
14	2	0	0	1	0	N
15	2	0	0	0	1	Y
16	2	0	0	0	0	N

transition probabilities can be obtained from the econometric models that are estimated. These probabilities result from the predicted purchase probabilities for each service with only one exception: once a customer stopped purchasing services from the company, it is assumed he is lost for good, so he remains in the inactive state with probability one.

Once an estimate of the probabilities in the transition matrix is obtained, calculation of CLV over a finite or infinite period is straightforward (Pfeifer and Carraway 2000). Let the Markov chain transition matrix P , which contains the probabilities p_{ij} with which a customer goes from state s_i to s_j in one period, be given by

$$\text{State at time } t \begin{pmatrix} s_1 \\ \vdots \\ s_K \end{pmatrix} \begin{pmatrix} \text{State at time } t+1 \\ p_{11} & \cdot & \cdot & p_{1K} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ p_{K1} & \cdot & \cdot & p_{KK} \end{pmatrix} = P \quad (5)$$

In particular, for a customer who is in state k at time t , the probability distribution for the subsequent period's states is given by the k th row of the transition matrix P , which can be written as $e_k P$, with e_k a vector of zeroes with a 1 at the k th place. Now $e_k P$ is a vector with the state probabilities in period $t+1$. These can be used to compute the state probabilities in period $t+2$, which are given by $e_k P P = e_k P^2$. More general, one can show that the state probabilities in year $t+\tau$ are given by $e_k P^\tau$.

Let *Statemargins* denote the vector stacking the annual customer profitabilities of each state. As only the services purchased are relevant for profits, the statemargins are the margins that correspond to the portfolio of services that is purchased in each state.⁵ Finally, we use $s_{i,t}$ to denote the state vector for customer i at time t , which is

⁵As in the other models, margins per service are constant over time.

a vector of zeros with a 1 at the position of the state for customer i at time t . With this notation, expected customer profitability in period $t+\tau$ is given by

$$E_t\{\text{Profit}_{i,t+\tau}\} = s_{i,t}P^\tau \text{Statemargins} \quad (6)$$

Predicted CLV levels are now easily obtained by summing up the appropriately discounted predicted future profit levels.

3 Data

Our application concerns the insurance industry. We have 6 years of behavioral data for 30,000 customers of a large Dutch direct writer. This direct writer sells many different types of insurance, some of them with very low ownership rates. We decided to include insurance types with ownership rates above 1% (see Table 3). The ownership of the other types of insurance is grouped in a category “other.” This is done because model estimation with ownership rates below 1% and stable behavior is troublesome, even on a sample of 30,000 customers (Donkers et al. 2003).

The time period considered in this database begins January 1, 1998 and ends January 1, 2003. All customers are active on January 1, 1998 and/or on January 1, 1999. Each year, the following information is observed: types of insurance purchased, start-date of the relationship, timing of last additional insurance purchase, timing of last cancellation of purchase of insurance, and membership of a loyalty program. Based on these data and information on the margins per insurance type, provided to us by the actuarial department of the insurance company, we can derive the following set of RFM variables: purchase recency dummy, and a cancellation

Table 3 Descriptive statistics ($n=24,055$, only customers active 1/1/1998)

Variable	January 1, 1998	January 1, 2003	
		All customers	Retained customers
Purchase rates			
Insurance 1, liability	31.9%	30.7%	36.2%
Insurance 2, car	40.2%	42.6%	50.2%
Insurance 3, furniture	31.5%	29.8%	35.2%
Insurance 4, house	22.9%	22.6%	26.7%
Insurance 5, legal aid	9.4%	11.5%	13.6%
Insurance 6, other	7.9%	7.2%	8.5%
Insurance 7, health	20.7%	24.5%	28.9%
Insurance 8, boat	4.1%	3.8%	4.5%
Insurance 9, caravan	5.5%	5.4%	6.4%
Insurance 10, life	15.5%	14.3%	16.9%
Insurance 11, motor	2.4%	2.9%	3.4%
Insurance 12, travel	3.4%	10.2%	12.0%
Purchase in last year	27.4%	6.1%	7.2%
Cancellation in last year	3.7%	5.4%	3.8%
Number of insurances	2.0	2.1	2.4
Indexed profit (1998=100)	100.0	109.7	129.4
Relationship age in years	8.2	—	13.2
Membership of loyalty program	25.1%	23.3%	27.5%

Table 4 Correlation matrix of insurance ownership in 1998 and 2003

		Insurance type 2003											
		1	2	3	4	5	6	7	8	9	10	11	12
Insurance type 1998	1	0.77**	0.05**	0.55**	0.34**	0.19**	0.06**	-0.06**	-0.01	0.04**	0.02*	0.01	0.09**
	2	-0.02*	0.61**	-0.02*	-0.04**	0.00	-0.01	-0.14**	-0.02*	0.05**	-0.08**	0.02*	0.05**
	3	0.60**	0.06**	0.79**	0.43**	0.21**	0.07**	-0.06**	-0.01	0.04**	0.03*	0.01	0.10**
	4	0.46**	0.08**	0.54**	0.83**	0.24**	0.05**	-0.02*	0.00	0.05**	0.08**	0.02*	0.11**
	5	0.45**	0.11**	0.44**	0.44**	0.83**	0.07**	0.01	0.00	0.04**	0.07**	0.02*	0.12**
	6	0.14**	-0.03*	0.13**	0.05**	0.06**	0.68**	-0.02*	-0.01	0.02*	0.04**	0.00	0.05**
	7	-0.14**	-0.17**	-0.14**	-0.10**	-0.04**	-0.01	0.75**	-0.02*	-0.01	0.07**	0.00	0.03**
	8	-0.10**	-0.10**	-0.10**	-0.06**	-0.03**	-0.03*	-0.10**	0.68**	0.01	-0.07**	0.00	0.00
	9	0.15**	0.23**	0.14**	0.12**	0.07**	0.02*	-0.03*	0.01	0.60**	0.01	0.02*	0.19**
	10	0.03*	-0.10**	0.05**	0.09**	0.04**	0.03*	-0.07**	-0.02*	0.01	0.79**	-0.01	0.02*
	11	0.00	0.15**	-0.02*	-0.01	0.03*	0.01	-0.05**	0.00	0.02*	-0.04**	0.46**	0.04**
	12	0.16**	0.15**	0.17**	0.13**	0.13**	0.05**	0.10**	0.00	0.11**	0.00	0.03*	0.77**

** : $p < 0.01$, * : $p < 0.05$

recency dummy, ownership dummies for each service, customer profitability, and relationship age. In Table 3, we report descriptive statistics on the type of insurance purchased and on the other explanatory variables for the beginning of our time period and the end of the period.

Due to customer defection, one would expect purchase rates of the insurances to decrease over time for the current customer base. However, the firm is relatively successful at cross-selling, and, in fact, average purchase incidence increased for many of the insurance types we consider. For the retained customers only, purchase incidence increased for all services. This is expected, since customers who purchase more services are more likely to stay with the firm. Average profits for all customers grew to 109.7 (1998=100), while the average profit for retained customers in 2003 equals 129.4, an increase of 29.4%. The retention rate in the year used for model estimation is 95.7%. The retention rate in the four subsequent years is 95.7, 96.2, 97.5, and 97.8%. In developing our model we have been concerned with dependencies across the purchase decisions. The existence of such dependencies becomes clear from Table 4, where we report Pearson correlation coefficients between ownership at 1998 and ownership at 2003. There are high correlations on the diagonal, so having a certain insurance type at 1998 is highly correlated with having that insurance type in 2003. More important, there are also substantial correlations across different insurance types.

4 Analysis

4.1 Time horizon of prediction

We use the first year of our data to estimate our models and to determine the aggregate and segment-level retention rates. These models are estimated for all customers active on January 1, 1998, where we model their behavior on January 1, 1999. The results are used to predict CLV for all customers active on January 1, 1999. In these predictions, we do not include customers defecting in the first year of data (1998–1999), while we do include customers acquired during that first year. We use the data available in the database on January 1, 1999 as predictors of future profitability. CLV is then calculated from the purchases in the subsequent 4 years (January 2000 till January 2003), with profits discounted at an annual rate of 10%. Although, theoretically, CLV predictions concern an infinite time horizon most firms use finite time horizons in practice. The time horizon of 4 years for CLV prediction is in line with recommendations of Rust et al. (2000). Note that in contrast to Eq. 1, we do not include profits for 1999 in the CLV calculation, as these profits are already observable. Including these profits would only inflate the predictive performance of all models.

4.2 Predictive performance measures

The out-of-sample predictive performance of the models is assessed with respect to three different tasks: (1) predicting the level of individual CLV, (2) predicting the ordering of customers based on CLV, and (3) valuing the total customer base. Correct predictions for the level of CLV are relevant when a company wants to target

customers with a CLV above a certain level, for example, because for customers with a lower CLV the marketing action will not be profitable. More often, however, companies will be interested in targeting their most profitable customers, without being interested in the precise level of CLV. To know how well a model selects the most profitable customers, an ordering-based measure of predictive performance should be used. Given current marketing practice, the ordering-based measure will be more relevant, since CLV is more often used as a segmentation device than as a device to manage profitability of marketing activities at the individual level (Zeithaml et al. 2001). Finally, correct prediction of the total value of the customer base is of interest for firm valuation purposes (Gupta et al. 2004).

To analyze how well each model predicts the level of CLV for each individual, we use the Mean Absolute Error (MAE) and the Root Mean Squared Error (RMSE) criteria (Leeflang et al. 2000). To facilitate their interpretation, we divided both measures by the average CLV and multiplied them by 100. In this way, they can be read as MAE/RMSE as percentage of average CLV. We also measure the predictive performance using a hit-rate criterion. For this hit-rate, we categorize all customers based on their true CLV value into four equal-sized groups with increasing levels of CLV. The hit-rate is then computed as the percentage of customers whose predicted CLV falls into the same category as their actual CLV. Suppose, for example, that the 25% most profitable customers (with the highest CLV) have a CLV of more than 200. The hit rate then measures how many of these customers also have a predicted CLV of more than 200.

The predictive performance with respect to the ordering of the customers is also evaluated with a hit-rate measure, see also Malthouse and Blattberg (2005). In contrast to the hit-rate for levels, this hit-rate does not consider the level of CLV, but only the ordering of the customers with respect to CLV. For the example above, the ordering-based hit-rate measures how many customers with an actual CLV above 200—the top 25% based on actual CLV—have a predicted CLV that is in the top 25% of predicted CLVs.

To measure predictive performance at the customer base level, based on individual-level CLV predictions (Gupta et al. 2004), we calculate the percentage deviation from the true value of the total customer base.

5 Empirical results

5.1 Estimation results

The estimation results of the Probit-model for retention are reported in the first column of Table 5. All variables have the expected signs: owning a type of insurance, membership of the loyalty program, and having purchased an additional service in the last year make a customer more likely to stay. Lower retention rates are predicted for customers who cancelled an insurance policy in the last year. The second column of Table 5 presents the estimation results of the parametric hazard rate model while the final column presents the estimation results of the nonparametric hazard rate model, excluding the 39 estimates for the nonparametric hazard rate itself, corresponding to the 40 different relationship ages in our sample

Table 5 Estimation results Probit model for customer retention

Variable	Probit model	Parametric duration model	Nonparametric duration model
Intercept	1.170** (0.105)	-1.175** (0.1082)	-1.702** (0.2131)
Ins 1	0.596** (0.051)	-0.8476** (0.108)	-0.9053** (0.1094)
Ins 2	0.622** (0.052)	-0.848** (0.1031)	-0.907** (0.1055)
Ins 3	0.478** (0.051)	-0.7142** (0.107)	-0.7445** (0.1079)
Ins 4	0.728** (0.061)	-1.4117** (0.1411)	-1.4858** (0.1438)
Ins 5	0.469** (0.067)	-1.4177** (0.2278)	-1.4802** (0.2289)
Ins 6	0.283** (0.042)	-0.7298** (0.1479)	-0.7917** (0.15)
Ins 7	0.626** (0.051)	-1.2127** (0.1227)	-1.2556** (0.1245)
Ins 8	0.298** (0.038)	-1.1579** (0.1862)	-1.1948** (0.1857)
Ins 9	0.210** (0.042)	-0.5089** (0.1778)	-0.5728** (0.1786)
Ins 10	0.742** (0.055)	-1.7774** (0.1466)	-1.8456** (0.1493)
Ins 11	0.153** (0.036)	-0.5392** (0.2183)	-0.7262** (0.2296)
Ins 12	0.231** (0.053)	-0.8637** (0.273)	-1.0921** (0.2928)
Loyalty program	0.221** (0.053)	-0.444** (0.1219)	-0.4009** (0.1208)
Purchase last year	0.184* (0.094)	-1.0378** (0.1073)	-0.5021** (0.1942)
Cancellation last year	-0.114** (0.028)	0.6862** (0.1425)	0.6679** (0.1423)
Relationship age <1	0.124 (0.087)	—	—
Relationship age <2	-0.065** (0.028)	—	—
γ_1	—	-0.0539** (0.0145)	—
γ_2	—	0.0014** (0.001)	—

Standard errors in parentheses

**: $p < 0.01$, *: $p < 0.05$

(and one parameter excluded for normalization). With so many parameters, the model might be somewhat over-parameterized and indeed the BIC selects the more parsimonious parametric hazard rate model. Moreover, the parametric hazard model is also preferred over the Probit model according to the BIC criterion. Note that this concerns model selection based on model fit, not based on predictive performance. The effects of the explanatory variables are very similar in the two proportional hazard models. Moreover, they are also similar to those in the choice model, except for the reversed sign and different scaling, which is as expected. This is in line with the result from Meyer (1990) that duration models with discrete time observations are equivalent to choice models.

The results from the multivariate Probit model are presented in Appendix B. All coefficients have the expected sign, at least when they are significantly different from zero. Purchasing a particular service makes a customer much more likely to repurchase that service in the next year. The loyalty program induces customers to buy more services and the recency variables seem to signal a change in the attitude of the customer. Customers who recently bought an additional service are more likely to buy even more, while the opposite holds for customers who recently cancelled an insurance policy. Also customers that are relatively new to the company are more likely to (continue to) buy services. The cross effects at the insurance level indicate that a number of insurances are grouped together. In particular, house, furniture, and liability insurances (insurance types 1, 3, and 4) are sold both individually and as a package. This is also confirmed by the estimated factor loading matrix, after varimax rotation, indicating that the purchase decisions on house (4), furniture (3) and liability (1) are highly correlated.

Table 6 Predictive performance at individual and customer base level

	Individual CLV levels			Individual CLV ordering	Customer base
	MAE (%) ^a	RMSE (%) ^a	Hit rate (%)	Hit rate (%)	% deviation
1. Relationship-level models					
1.1. Status quo	19.4	40.8	45.1	78.4	-4.3
1.2. Profit regression	23.1	40.4	54.7	78.4	-1.2
1.3. Retention (segmented)					
Aggregate retention rate	23.9	41.9	77.0	78.4	-11.3
Segmented retention rate					
Recency of last purchase/ cancellation	25.1	42.5	76.9	72.8	-13.3
Frequency (number of insurances)	25.2	42.0	78.3	75.9	-12.7
Monetary value (high vs. low profit)	24.1	42.0	77.2	78.4	-11.6
RFM-index (high vs. low)	24.6	41.9	77.4	71.8	-12.2
1.4. Probit model	24.6	41.8	78.4	71.3	-11.9
1.5. Bagging	25.2	41.9	77.6	70.3	-13.3
1.6. Duration model					
Parametric hazard rate	24.7	41.9	78.4	69.8	-12.1
Non-parametric hazard rate	25.1	42.0	78.3	71.7	-12.9
1.7. Tobit II model	21.1	40.9	78.3	78.4	-6.3
2. Service-level models					
2.1 Independent choice models	24.8	40.1	69.0	71.1	-4.9
2.2 Independent duration models	31.0	46.0	75.1	67.5	-21.5
2.3 Multivariate probit model	24.9	40.2	69.7	71.7	-6.3
2.4 Multivariate duration model	31.8	46.8	74.3	65.8	-22.3

^a Values are measured as % of the average CLV

For the service-level duration models, most effects are also as expected. The most remarkable finding, however, is that compared to the service-level choice models fewer effects are statistically significant. This has a natural explanation, as the model uses less information than the service-level choice models. The duration model for a service is estimated only for those customers currently purchasing that service. The parameter estimates are thus based on much smaller samples than those of the choice models. Moreover, choice models determine purchase probabilities using information on the differences between the customers who do and those who don't buy the service and on the customers that start or stop buying a service. The duration model concerns the process of continuing buying a service, so it can only use the information on customers currently purchasing that service. Possibly for the same reason, the model with a single factor is selected, based on the BIC criterion,⁶ which is extremely similar to the first factor in the service-level choice models. As the two models are rather different and estimated independently, this provides clear evidence that the factor based approach advocated by Kamakura et al. (2003) is able to detect dependencies between different model equations.

⁶The AIC criterion selected the same number of factors as the BIC criterion.

5.2 Predictive performance

The predictive performance of our models is presented in Table 6. We start with a discussion of the predictive performance for individual CLV levels in the first three columns. The first remarkable finding is the large difference in performance across the three measures we consider. The second result is the poor performance of the service-level duration model, no matter what criterion is used. This poor performance could be due to the type of information that is inferred from and the limited use of the data in model estimation, discussed in the previous section.

The best model for CLV level prediction, according to MAE, is the status quo model followed by the Tobit II model; for RMSE, it is the multivariate Probit model with either dependent or independent purchase decisions. According to the hit-rate criterion, the Tobit II model is ranked high, as well as the models accounting for retention, including the recently advanced bagging technique. That the status quo model does a particularly good job when the MAE is used is due to the relatively large number of customers that indeed do not change their purchase behavior over time. For these customers, the prediction error is zero. In contrast to this, the model makes large prediction errors for customers that defect or cross-buy. Most other models will not perfectly predict CLV for the customers that do not change their portfolio, but, in general, smaller errors are made for the customers that defect. Compared to MAE, RMSE punishes large mistakes much more. Our conclusion is that, in comparison with the status quo model and the retention-based CLV models, the multivariate Probit model makes more mistakes, though fewer big mistakes. Given the large differences across the measures, it is not clear what type of model should be favored, with the exception of the service-level duration models, because these models perform poorly on all criteria. Which model performs best depends on what type of prediction error is deemed important by the company.⁷

None of the models predicts customer-level CLV perfectly, as could be expected, but are the errors made by the models acceptable? Given the discrete nature of the purchase process for insurances—one either does or does not buy a particular type of insurance—large prediction errors should be expected. In comparison with predicting the overall mean, all models reduce MAE by almost 60% and RMSE by more than 40%. The models are thus clearly able to use customer specific information to improve customer-level CLV predictions, which is further illustrated by the hit rate that exceeds the 25% benchmark of random predictions by far.

In cases where only the ordering of customers is important, the status quo model performs very well. Its performance is the same as that of the profit regression model, the model accounting for average retention, or for retention based on monetary value, and the Tobit II model. Here it is important to realize that all these orderings are identical, so the status quo model might be favored because of its simplicity. Most models are better able to predict the ordering of the customers than

⁷A company might redo the analysis in our paper with its own loss function to see which model performs best in that particular case.

the actual level of CLV, with the exceptions being the models accounting for retention only and the service-level duration models.

When using CLV models to value the whole customer base, all models predict a value that is below the actual value. This even holds for the status quo model, which would usually overvalue, as it does not account for customer defection. These predictions below the actual value of the customer base are due to the company's success in cross-selling activities to the existing customers. Indeed, about 12% of the company's total CLV is due to services customers did not buy at the time the CLV predictions are made. The best performance is shown by the profit regression model that accounts for this growth in customer profitability. In contrast, the retention-based models all attribute a similar value to the customer base, which is about 12% below its actual value, which is quite substantial. This large difference indicates that ignoring the potential value embedded in cross-selling in multi-service industries can substantially harm the predictive performance. The models that take customer retention and cross-buying into account separately and then combine the two types of behavior into a single prediction of CLV have a performance that is in between the status quo model and the profit regression model—which do very well—and the models accounting for retention only. The amount of undervaluation of the customer base is 6.3% for the Tobit II model and for the multivariate Probit model with dependent purchase decisions. The Probit model with independent purchase decisions does even better with an underestimation of total CLV with only 4.9%. Here the service-level duration models perform really poorly. An explanation is that these models not only ignore the value of cross-buying, but also incorporate the indirect effects of changes in ones portfolio. Even though this is conceptually valid, it increases the consequences of ignoring cross-buying, as the effect of larger portfolios on future purchase behavior is not captured.

Thus, for the valuation of the total customer base in multi-service industries, it seems to be important to model not only retention, but also cross-buying. The good performance of the status quo model might have little generalizability. In our application the effects of customer defection and cross-buying are similar in size and therefore the net effect is small. In other markets, one of the effects is likely to dominate the other, resulting in larger prediction errors for the status quo model. A good alternative could be the profit regression model, which will pick up the net effect of cross-buying and defection. Also the models that model cross-buying and retention separately, i.e. the service-level choice models and the Tobit II model, perform rather well.

An interesting question is whether the predictive performance of the models varies over time. For example, one might expect the status quo model to predict extremely well in the short run, but worse in the long run. For the more complex models, this deterioration is expected to be smaller as these models allow for changes in behavior that are more likely to occur as time passes by. To investigate this phenomenon, we consider the predictive performance of all models for the profits earned in 2000 and in 2003, the first and last (fourth) year of our prediction period. The results are presented in Table 7. The surprising result is that the relative position for most models is the same for the periods considered, with obviously the

Table 7 Predictive performance at individual and customer base level, First and fourth year

	Individual CLV levels			Individual CLV ordering	Customer base
	MAE	RMSE	Hit rate (%)	Hit rate (%)	% deviation
Value in 2000 (first year)					
1. Relationship-level models					
1.1. Status quo	10.7	33.3	89.7	89.7	-2.4
1.2. Profit regression	13.1	33.2	61.4	89.7	-1.0
1.3. Retention (segmented)					
Aggregate retention rate	14.4	33.8	88.3	89.7	-6.6
Segmented retention rate					
Recency of last purchase/ cancellation	14.2	33.8	88.6	83.2	-6.4
Frequency (number of insurances)	14.1	33.7	89.7	86.3	-6.2
Monetary value (high vs. low profit)	13.6	33.6	89.6	89.7	-5.6
RFM-index (high vs. low)	13.9	33.6	89.7	81.1	-6.0
1.4. Probit model	13.8	33.6	89.7	80.8	-5.9
Bagging	14.2	33.6	89.3	78.9	-6.4
1.5. Duration model					
Parametric hazard rate	13.9	33.6	89.7	78.6	-6.0
Non-parametric hazard rate	14.2	33.6	89.7	80.9	-6.3
1.6. Tobit II model	13.3	33.2	86.5	89.7	0.2
2. Service-level models					
2.1 Independent choice models	14.7	33.0	71.5	79.0	-2.0
2.2 Independent duration models	18.4	35.5	87.8	78.8	-11.3
2.3 Multivariate Probit model	14.8	33.0	76.6	79.1	-2.8
2.4 Multivariate duration model	18.9	35.8	87.2	74.8	-11.7
Value in 2003 (fourth year)					
1. Relationship-level models					
1.1. Status quo	28.1	55.3	73.6	73.6	-6.1
1.2. Profit regression	34.0	54.7	42.5	73.6	-1.1
1.3. Retention (segmented)					
Aggregate retention rate	30.7	55.6	72.4	73.6	-10.2
Segmented retention rate					
Recency of last purchase/ cancellation	37.3	58.4	71.3	68.1	-20.5
Frequency (number of insurances)	37.6	57.6	70.9	70.1	-19.5
Monetary value (high vs. low profit)	35.9	57.4	72.0	73.6	-17.9
RFM-index (high vs. low)	36.6	57.4	71.1	66.8	-18.8
1.4. Probit model	36.5	57.2	72.5	65.6	-18.3
Bagging	37.5	57.5	71.0	65.7	-20.5
1.5. Duration model					
Parametric hazard rate	36.7	57.4	72.7	64.6	-18.4
Non-parametric hazard rate	37.3	57.6	72.4	65.7	-19.7
1.6. Tobit II model	32.9	56.3	72.3	73.6	-13.0
2. Service-level models					
2.1 Independent choice models	37.3	54.8	58.1	59.1	-8.6
2.2 Independent duration models	45.6	63.6	65.8	62.0	-32.0
2.3 Multivariate Probit model	37.3	54.9	60.3	65.8	-10.6
2.4 Multivariate duration model	46.7	64.8	65.4	66.0	-33.4

quality of the predictions in the first year being better than that of the last year. The only model that does seem to change its relative position is the Tobit II model which performs somewhat better in the first year than in the fourth year.

Summarizing our results, there is no “best” model for the prediction of individual CLV levels. For segmentation purposes, the simple status quo model performs best. The more complex models are only interesting for the valuation of the total customer base, but here the profit regression and Tobit II models are also good alternatives. Overall, the performance of the complex service-level models with or without dependence in the purchase decisions is somewhat disappointing, especially given the large efforts required to obtain parameter estimates and CLV predictions for these models.

6 Discussion

In this paper we used a large variety of models to predict CLV. The models varied substantially in their degree of complexity, ranging from a simple status quo model to complex service-level models. The main conclusions of our research are: (1) simple models perform well and (2) focusing only on customer retention is not sufficient; cross-buying needs to be accounted for.

The fact that simple models perform well will provide a comforting idea to practitioners who frequently use relatively simple models. In general, management is reluctant to adopt complex models in cases where they perform very well (Verhoef et al. 2003). Thus, for the purpose of predicting individual CLV, the adoption of complex models by practitioners should not be expected. We realize that this conclusion is based on a single case study.

The more complex models still might have added value over the simple models, as they can aid in the targeting of customers for marketing activities. Only the service-level choice models investigate the decisions to purchase each insurance type. Such a model can therefore provide more insights for targeting customers at the individual insurance level. For example, the multivariate Probit model provides the likelihood that a customer will purchase an additional service (see also Kamakura et al. 2003); information that cannot be provided by the simple models. All service-level models also provide an estimate of the probability that a customer cancels a certain type of insurance, providing information for targeted service-level retention campaigns.

The statement that simple models work well does not hold for the models that only deal with customer retention. These models systematically underestimate customer value, as they do not account for cross-buying and the corresponding growth in profits. As a result, the simple profit regression model, which can deal with growth rates in profits, outperforms these models. The model that has a good performance on all criteria is the Tobit II model, which combines a regression model for profit growth with a Probit model for customer retention, thereby capturing both dimensions of customer behavior. The multivariate Probit model also accounts for these behaviors, but concerns decisions at the service level. The reduced parsimony of the model—it has many more parameters—might have harmed its predictive performance.

The duration models have a decent performance when applied at the relationship level. The usefulness of duration models for modelling service-level behavior,

however, seems to be limited. A first problem is the difficulty of modelling cross-buying behaviour with a duration model. Another drawback is that the duration models ignore the information that exists in the difference between current buyers and non-buyers, while this information is available and used by the choice models. At the relationship level, the same information is used by the two approaches, i.e. both analyze which customer from the current database is most likely to end his/her relationship. Note that these problems arise from the multiple services setting of our application, not from the specifics of the contractual setting; this only required an adaptation of the estimation procedure.

We want to stress again that our results are only based on a single application of the models in a relatively stable industry. However, given the nascent nature of CLV research, our conclusions might resonate beyond this case study. Moreover, the importance of cross-buying suggests that customer behavior is not that stable. It is, however, important that future research replicates this study in different industries, leading to empirical generalizations on this topic.

The predicted CLV in our models can be used to segment customers. Using this segmentation firms can, for example, develop different service concepts for each segment. For example, high value customers are serviced rather fast and may choose the preferred channel. Low value customers receive lower service levels, while they are directed to low cost channels, such as the Internet (see also Malthouse and Blattberg (2005)). We recommend firms only to use these segmentations, in case CLV can be predicted with reasonably accuracy. Our study provides customer intelligence managers with several methods for CLV prediction, which can be tested in their market.

A next step in the use of CLV in marketing, is the development of an individual specific marketing mix that optimizes each customer's CLV (e.g. Rust and Verhoef 2005; Venkatesan and Kumar 2004). Current models aiming to do so, usually only consider relatively short time periods (i.e. 1 year). Developing these optimal marketing mixes for individual customers is probably rather difficult for longer time periods, as the optimal marketing mix for the next few years will be highly dependent on environmental factors, such as the competitive setting and changes in regulation. So far, researchers have not touched upon the issue. This topic goes well beyond the scope of this study. It is certainly a very interesting avenue for future research.

Acknowledgements We acknowledge the data-support of a Dutch financial service provider. We thank Aurelie Lemmens for providing us the bagging procedure and support with the implementation. We thank Jan-Benedict Steenkamp and seminar participants at the Catholic University of Leuven, the University of Groningen and the University of Münster for their helpful comments. We also acknowledge the helpful comments of the editor and two anonymous reviewers.

Appendix A

In this appendix, we more elaborately describe the specification and estimation of the various models in Table 1. We start with a description of the relationship-level models, and then discuss the service-level models.

A.1 Relationship-level models

The profit regression model (1.2), as discussed by Malthouse and Blattberg (2005), is based on the following linear regression model:

$$\text{Profit}_{t+1} = \alpha_0 + \alpha_1 \text{Profit}_t + \varepsilon_t \quad (7)$$

This model is estimated using Ordinary Least Squares, but more advanced approaches could be used, see Malthouse and Blattberg (2005).

The customer retention models (1.3 and 1.4) are nested within the Probit model defined by

$$\text{Ret}_{t+1}^* = x' \beta + \varepsilon_t \quad (8)$$

$$\text{Ret}_{t+1} = 1(\text{Ret}_{t+1}^* > 0) \quad (9)$$

In these equations, Ret_{t+1}^* is a continuous latent variable, and the observed retention is obtained by dichotomizing the latent variable. The resulting probability that a customer is retained is then the well-known formula $P(\text{Ret}_{t+1}=1 | x) = \Phi(x' \beta)$. The model parameters can be estimated by maximum likelihood procedures. The customer retention models 1.3 and 1.4 differ in the explanatory variables included. The segment-level retention rates result from the Probit model with segment membership dummies as explanatory variables. The model referred to as the Probit model contains all explanatory variables, including two dummies for the short relationship lengths.

The bagging method (1.5) is based on classification trees, but improves this procedure by averaging the predictions using classification trees that are built on random samples (with replacement) from the population. As we aim for CLV prediction and not customer classification, we use the average predicted probability for each customer across all samples as an estimate of the customer's retention probability. Using the average classification is also not successful as all customers are always classified as loyal due to the low incidence of customer defection. For a detailed description of the bagging procedure the reader is referred to Lemmens and Croux (2006) or Breiman (1996).

The relationship-level duration models (1.6) are based on the proportional hazard model where the hazard rate is given by $h(t, x) = \exp(x' \beta) h(t)$, where relationship duration is not included as an explanatory variable in x . The effect of the explanatory variables is a proportional shift in the hazard rate, which itself only depends on time t , the time since the start of the relationship. For a detailed discussion of hazard rate models we refer the reader to Helsen and Schmittlein (1993). Here, we present the most important results needed for model estimation. For our purposes, the most important feature of duration models is the survival function $S(t, x)$, which is the probability that a customer's relationship exceeds t before it ends. When modeling customer retention, we are interested in the probability that the customer's relationship will last for another year, at least in the contractual setting under investigation. This is equivalent to no critical events happening in the year the current contract runs, leading to renewal of the contract. This probability of a relationship lasting for another year, given that the relationship already lasted for t years, equals $S(t+1, x)/S(t, x)$; the probability of the relationship length exceeding $t+1$ periods, conditional on knowing that it already exceeded t periods. The survival function $S(t, x)$ equals $\exp(-\int_0^t h(\tau, x) d\tau)$, see Helsen and Schmittlein. The

corresponding retention probability, cf. Meyer (1990), then equals $\exp(-\int_t^{t+1} h(\tau, x) d\tau)$ or after filling out the proportional hazard specification

$$\begin{aligned} P(\text{Ret}_{t+1} = 1 | x, \text{relationship duration} = t) \\ = \exp\left(-\exp(x'\beta) \int_t^{t+1} h(\tau, x) d\tau\right) \end{aligned} \quad (10)$$

Note that with the hazard rate being non-negative, this indeed is a probability between zero and one.

For the hazard rate we implement the flexible quadratic Box–Cox formulation proposed by Helsen and Schmittlein (1993), which is given by $h(t) = \exp(\gamma_0 + \gamma_1(t-1) + \gamma_2(t^2-1)/2)$. As an alternative, we also implemented a model that is completely non-parametric in the hazard rate, given our discrete observation scheme. It is implemented by having the hazard rate constant within each year and the level being estimated by a dummy variable for each year, yielding: $h(t) = \sum_{j \geq 0} I(j \leq t < j+1) \exp(\gamma_j)$, with $I(j \leq t < j+1)$ the indicator function having the value 1 in case t is between j and $j+1$ and 0 otherwise. The hazard models are estimated with maximum likelihood, using the formula for the probability of retention as a function of the parameters to describe customer retention. As a constant is included in x , we set $\gamma_0 = 0$ for both models as a normalization.

The *Tobit II model* (1.7) combines customer retention and profit regression into a single model. More precisely, it combines Eqs. 7–9 and allows for correlation between the errors. The major difference with the profit regression model is that it only uses Eq. 7 to predict profits conditional on retention, see Franses and Paap (2001) for more details. The model is estimated with maximum likelihood with past profits as explanatory variables in the profit Eq. 7 and the full set of explanatory variables in the retention Eq. 8.

Note that the explanatory variables used to predict retention and those used to predict profits are different. To predict future profits, only past profits and an intercept are included, while all types of variables are included in the retention models. The reason for this is that effects in the retention equation imply a time invariant effect on the customer retention probability. In the profit equation, they imply a constant growth factor, as past profits are included. Such a constant growth factor is unlikely to hold for the types of variables we include.

A.2 Service-level models

For the models without dependence between the services (models 2.1 and 2.2), model estimation at the service level is equivalent to model estimation at the relationship-level and needs no further discussion. The most important aspect in our specification is that the parameters are allowed to vary across the services, so each service has its own set of parameters.

Estimation of the Probit and duration models with dependence among the service purchase decisions (models 2.3 and 2.4) is more complex. The dependencies for both models are assumed to follow a factor structure, building on the ideas in Kamakura et al. (2003). More specifically, the dependencies across the decisions on

the J services result from an unobserved component ε that is added to the systematic component, $x'\beta$, for both the Probit models (Eq. 8) and the duration models (Eq. 10). This unobserved component is characterized by a factor structure: $(\varepsilon_{i1,t}, \dots, \varepsilon_{iJ,t})^T = \mathbf{A}\mathbf{f}_{i,t}$. Here, $\mathbf{f}_{i,t}$ denotes the vector of unobserved factor scores with a $N(0,1)$ distribution and \mathbf{A} the matrix with factor loadings. The model is estimated using the method of smooth simulated maximum likelihood (Kamakura et al. 2003; Stern 1992). This procedure starts with a large number of random draws, R , from the assumed distribution of the factor scores $\mathbf{f}_{i,t}$. For each draw the likelihood is computed and subsequently averaged over the draws. The likelihood is then approximated by

$$\frac{1}{R} \sum_{r=1}^R L(y|x, \mathbf{f}_r). \quad (11)$$

This approximation error is negligible when the number of draws R grows with N (Stern 1997).

The imposed factor structure on the unobserved components and the assumption of independence conditional on the factors and explanatory variables yields a very attractive estimation procedure by implementing the ideas of Stern (1992). Its simplicity arises from the conditional independence, making $L(y|x, \mathbf{f}_r)$ the product of the likelihoods per service, which is easy to obtain. The number of factors to be included in the model is determined by minimizing the BIC criterion. For the multivariate Probit models, this resulted in two factors being included, and one factor is included to model the dependence in the duration models.

A.3 Model prediction

For most models, prediction of customer behavior in the next period is straightforward. The estimated parameter values can be substituted in the model Eqs. 7, 9, or 10. The same holds for the service-level models without cross-equation dependence, where purchase probabilities for each service can be obtained straightforwardly by substituting the service-specific parameters estimates in the model equations. The probability of purchasing a certain portfolio of services is then simply the product of purchasing each individual service.

For the multivariate service-level models, obtaining purchase probabilities for all possible portfolios of services is more complicated. Here, we rely again on the simulation approach used also for model estimation. Given the factor scores it is straightforward to obtain the probability of purchasing a portfolio of services, as these are independent conditional on the factor scores. The unconditional purchase probability for a given portfolio of services is approximated by averaging the purchase probabilities given the factor scores over a large number of draws from the distribution of the factors scores. The result is an approximation of the probability of purchasing a portfolio of services that is also used in estimation, see Eq. 11.

Appendix B

Table 8.

Table 8 Estimation results for the multivariate Probit model

Variable	Ins 1	Ins 2	Ins 3	Ins 4	Ins 5	Ins 6	Ins 7	Ins 8	Ins 9	Ins 10	Ins 11	Ins 12
Intercept	-7.641** (0.521)	-2.014** (0.039)	-7.539** (0.460)	-4.051** (0.125)	-5.472** (0.285)	-2.944** (0.069)	-2.207** (0.039)	-3.023** (0.130)	-2.974** (0.109)	-2.551** (0.055)	-2.737** (0.073)	-2.563** (0.061)
Ins 1	12.380** (0.843)	0.026 (0.044)	0.564** (0.130)	0.315** (0.082)	-0.123 (0.109)	0.050 (0.078)	0.089* (0.050)	0.011 (0.132)	0.108 (0.087)	0.045 (0.077)	0.015 (0.109)	0.063 (0.069)
Ins 2	0.344** (0.121)	3.350** (0.044)	0.404** (0.122)	0.157* (0.075)	0.328** (0.101)	0.146* (0.063)	-0.033 (0.039)	-0.215** (0.088)	0.206** (0.065)	-0.038 (0.055)	0.268** (0.069)	0.107* (0.049)
Ins 3	0.739** (0.134)	0.051 (0.047)	12.010** (0.742)	0.295** (0.080)	0.244* (0.110)	0.202** (0.080)	0.010 (0.054)	-0.030 (0.128)	-0.020 (0.100)	0.021 (0.084)	-0.042 (0.111)	0.058 (0.075)
Ins 4	1.151** (0.147)	0.108** (0.041)	1.326** (0.144)	6.348** (0.155)	0.532** (0.103)	0.151* (0.072)	0.058 (0.048)	0.037 (0.143)	0.063 (0.092)	-0.084 (0.066)	0.155* (0.086)	-0.015 (0.062)
Ins 5	0.428* (0.222)	0.089 (0.049)	0.330 (0.244)	0.252* (0.123)	8.989** (0.439)	0.184* (0.082)	0.087 (0.055)	0.180 (0.155)	0.127 (0.091)	0.176* (0.078)	0.134 (0.090)	0.155** (0.064)
Ins 6	0.033 (0.174)	-0.046 (0.055)	0.013 (0.178)	-0.059 (0.125)	-0.198 (0.188)	4.118** (0.070)	0.027 (0.058)	0.027 (0.126)	-0.172 (0.104)	0.129* (0.075)	0.036 (0.104)	-0.086 (0.081)
Ins 7	0.204 (0.154)	0.053 (0.041)	0.229 (0.160)	0.060 (0.099)	0.179 (0.130)	0.224** (0.066)	3.586** (0.043)	0.009 (0.115)	0.008 (0.083)	0.122* (0.063)	-0.075 (0.084)	0.330** (0.054)
Ins 8	-0.616* (0.367)	0.021 (0.086)	-0.148 (0.280)	-0.128 (0.196)	-0.245 (0.291)	-0.028 (0.148)	-0.034 (0.099)	4.430** (0.171)	0.271* (0.152)	-0.070 (0.117)	0.084 (0.142)	-0.007 (0.131)
Ins 9	-0.483* (0.263)	0.067 (0.065)	-0.333 (0.272)	-0.191 (0.119)	-0.154 (0.205)	-0.142 (0.117)	0.017 (0.077)	-0.200 (0.169)	3.943** (0.126)	0.037 (0.091)	0.045 (0.106)	0.258** (0.073)

Table 8 (continued)

Variable	Ins 1	Ins 2	Ins 3	Ins 4	Ins 5	Ins 6	Ins 7	Ins 8	Ins 9	Ins 10	Ins 11	Ins 12
Ins 10	-0.092 (0.150)	-0.096* (0.046)	-0.077 (0.159)	-0.079 (0.115)	-0.111 (0.131)	0.185* (0.079)	0.100* (0.046)	-0.282* (0.125)	0.079 (0.081)	4.167** (0.057)	-0.140 (0.091)	-0.102 (0.065)
Ins 11	0.482 (0.393)	-0.050 (0.093)	0.398 (0.366)	-0.021 (0.269)	0.506* (0.260)	0.057 (0.195)	0.105 (0.118)	0.131 (0.182)	-0.029 (0.187)	-0.027 (0.149)	3.740** (0.083)	0.198* (0.112)
Ins 12	0.052 (0.343)	0.146** (0.071)	0.154 (0.309)	0.231 (0.169)	0.313 (0.190)	-0.213 (0.143)	0.227** (0.074)	-0.070 (0.223)	0.194 (0.123)	0.025 (0.155)	0.292** (0.121)	3.868** (0.099)
Relationship age <1	-0.183 (0.250)	0.126* (0.061)	-0.005 (0.251)	0.071 (0.158)	-0.061 (0.206)	0.193* (0.117)	0.168** (0.069)	-0.190 (0.177)	-0.011 (0.148)	-0.207* (0.119)	0.012 (0.116)	-0.177* (0.095)
Relationship age <2	0.010 (0.235)	0.024 (0.054)	0.082 (0.202)	0.187 (0.164)	0.128 (0.176)	0.241** (0.093)	0.050 (0.060)	0.061 (0.154)	0.056 (0.167)	0.102 (0.092)	-0.072 (0.128)	-0.014 (0.100)
Loyalty program	1.535** (0.158)	0.323** (0.036)	1.276** (0.140)	0.479** (0.073)	0.895** (0.100)	0.236** (0.069)	0.225** (0.043)	0.358** (0.125)	0.270** (0.077)	0.212** (0.065)	0.173* (0.082)	0.421** (0.059)
Purchase last year	-0.142 (0.174)	0.153** (0.047)	-0.059 (0.180)	-0.048 (0.107)	-0.022 (0.135)	-0.112 (0.082)	0.205** (0.053)	0.254* (0.113)	0.019 (0.098)	0.087 (0.079)	0.123 (0.091)	0.058 (0.062)
Cancellation last year	-0.609** (0.232)	-0.074 (0.061)	-0.976** (0.222)	-0.167 (0.127)	-0.455* (0.216)	0.029 (0.112)	0.103 (0.082)	-0.028 (0.191)	0.154 (0.100)	-0.011 (0.108)	-0.083 (0.151)	-0.053 (0.100)

Standard errors in parentheses

**; $p < 0.01$, *; $p < 0.05$

References

- Berger, P. D., & Nasr, N. I. (1998). Customer lifetime value: Marketing models and applications. *Journal of Interactive Marketing*, 12(1), 17–30.
- Bolton, R. N. (1998). A dynamic model of the duration of the customer's relationship with a continuous service provider: The role of satisfaction. *Marketing Science*, 17(1), 45–65.
- Bolton, R. N., Kannan, P. K., & Bramlett, M. D. (2000). Implications of loyalty program membership and service experiences for customer retention and value. *Journal of the Academy of Marketing Science*, 28(1), 95–108.
- Bolton, R. N., Lemon, K. N., & Verhoef, P. C. (2004). The theoretical underpinnings of customer asset management in service industries. *Journal of the Academy of Marketing Science*, 32(3), 271–292.
- Breiman, L. (1996). Bagging predictors. *Machine Learning*, 24(2), 123–140.
- Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society. Series B*, 34, 187–220.
- Donkers, B., Franses, P. H., & Verhoef, P. C. (2003). Using selective sampling for binary choice models. *Journal of Marketing Research*, 40(4), 492–497.
- Fader, P. S., & Hardie, B. G. S. (2001). Forecasting repeat sales at CDNOW: A case study. *Interfaces*, 31, S94–S107 (May–June).
- Fader, P. S., Hardie, B. G. S., & Huang, C.-Y. (2004). A dynamic changepoint model for new product sales forecasting. *Marketing Science*, 23(1), 50–65.
- Franses, P. H., & Paap, R. (2001). *Quantitative models in marketing research*. Cambridge: Cambridge University Press.
- Gremler, D. D. (2004). The critical incident technique in service research. *Journal of Service Research*, 7(1), 65–89.
- Gupta, S., Lehmann, D. R., & Stuart, J. A. (2004). Valuing customers. *Journal of Marketing Research*, 41(1), 7–18.
- Hansotia, B., & Wang, P. (1997). Analytical challenges in customer acquisition. *Journal of Direct Marketing*, 11(2), 7–19.
- Helsen, K., & Schmittlein, D. C. (1993). Analyzing duration times in marketing: Evidence for the effectiveness of hazard rate models. *Marketing Science*, 12(4), 395–414.
- Hogan, J. E., Lemon, K. N., & Rust, R. T. (2002). Customer equity management: Charting new directions for the future of marketing. *Journal of Service Research*, 5(1), 4–12.
- Jain, D., & Singh, S. S. (2002). Customer lifetime value research in marketing: A review and future directions. *Journal of Interactive Marketing*, 16(2), 34–46.
- Jain, D., & Vilcassim, N. J. (1991). Investigating household purchase timing decisions: A conditional hazard function approach. *Marketing Science*, 10(1), 1–23.
- Kamakura, W. A., Kossar, B. S., Wedel, M. (2004). Identifying innovators for the cross-selling of new products. *Management Science*, 50(8), 1120–1133.
- Kamakura, W. A., Mela, C., Ansari, A., Fader, P., Iyengar, R., Naik, P., et al. (2005). Choice models and customer relationship management. *Marketing Letters*, 16(3), 192–279.
- Kamakura, W. A., Ramaswamy, S., & Srivastava, R. K. (1991). Applying latent trait analysis in the evaluation of prospects for cross selling of financial service. *International Journal of Research in Marketing*, 8, 329–349.
- Kamakura, W. A., Wedel, M., Rosa, F., & Mazzon, J. (2003). Cross-selling through database marketing: A mixed data factor analyzer for data augmentation and prediction. *International Journal of Research in Marketing*, 20(1), 45–65.
- Keaveney, S. M. (1995). Customer switching behavior in service industries: An exploratory study. *Journal of Marketing*, 59(April), 71–82.
- Knott, A., Hayes, A., & Neslin, S. A. (2002). Next-product-to-buy models for cross-selling applications. *Journal of Interactive Marketing*, 16(3), 59–75.
- Leeflang, P. S. H., Wittink, D. R., Wedel, M., & Naert, P. A. (2000). *Building models for marketing decisions*. Boston: Kluwer.
- Lemmens, A., & Croux, C. (2006). Bagging and boosting classification trees to predict churn. *Journal of Marketing Research*, 43(May), 276–286.
- Li, S., Sun, B., & Wilcox, R. T. (2005). Cross-selling sequentially ordered products: An application to consumer banking services. *Journal of Marketing Research*, 42(2), 233–239.
- Malthouse, E. C., & Blattberg, R. C. (2005). Can we predict customer lifetime value. *Journal of Interactive Marketing*, 19(1), 2–16.

- Manchanda, P., Ansari, A., & Gupta, S. (1999). The shopping basket: A model for multicategory purchase incidence decisions. *Marketing Science*, 18(2), 95–114.
- Meyer, B. D. (1990). Unemployment insurance and unemployment spells. *Econometrica*, 58(4), 757–782.
- MSI (2004). *Research priorities 2004–2006*. Boston.
- Pfeifer, P. E., & Carraway, R. L. (2000). Modeling customer relationships as Markov chains. *Journal of Interactive Marketing*, 14(2), 43–55.
- Prentice, R. L., & Gloeckler, L. A. (1978). Regression analysis of grouped survival data with application to breast cancer data. *Biometrics*, 34(1), 57–67.
- Reinartz, W., & Kumar, V. (2000). On the profitability of long-life customers in a non contractual setting: An empirical investigation and implications for marketing. *Journal of Marketing*, 64(4), 17–35.
- Rossi, P. E., McCulloch, R. E., & Allenby, G. M. (1996). The value of purchase history data in target marketing. *Marketing Science*, 15(4), 321–340.
- Rust, R. T., Lemon, K. N., & Zeithaml, V. A. (2004). Return on marketing: Using customer equity to focus marketing strategy. *Journal of Marketing*, 68(1), 109–127.
- Rust, R. T., & Verhoef, P. C. (2005). Optimizing the marketing mix interventions in intermediate-term CRM. *Marketing Science*, 24(3), 477–489.
- Rust, R. T., Zeithaml, V. A., & Lemon, K. N. (2000). *Driving customer equity: How customer lifetime value is reshaping corporate strategy*. New York: Free.
- Schmittlein, D. C., Morrison, D. G., & Colombo, R. (1987). Counting your customers: Who are they and what will they do next. *Management Science*, 33(1), 1–24.
- Stern, S. (1992). A method for smoothing simulated moments of discrete probabilities of multinomial probit models. *Econometrica*, 60(4), 943–952.
- Stern, S. (1997). Simulation-based estimation. *Journal of Economic Literature*, 35(4), 2006–2039.
- Venkatesan, R., & Kumar, V. (2004). A customer lifetime value framework for customer selection and resource allocation. *Journal of Marketing*, 68(4), 106–125.
- Verhoef, P. C., Spring, P. N., Hoekstra, J. C., & Leeflang, P. S. H. (2003). The commercial use of segmentation and predictive modeling techniques in database marketing in The Netherlands. *Decision Support Systems*, 34(4), 471–481.
- Zeithaml, V. A., Rust, R. T., & Lemon, K. N. (2001). The customer pyramid: Creating and serving profitable customers. *California Management Review*, 42(4), 118–142.