Forecasting customer behaviour in a multi-service financial organisation: a profitability perspective*

A. Audzeyeva[a,b], B. Summers[a] and K. R. Schenk-Hoppé[a,c]

[a] Leeds University Business School, University of Leeds; [b] National Australia Group Europe; [c] School of Mathematics, University of Leeds

Abstract

This paper proposes a novel approach to the modelling of Customer Lifetime Value, a metric indicating the profit-generating potential of customers, which provides a key business tool for the customer management process. Existing approaches are problematic due to the heterogeneous and also multidimensional nature of customer behaviour. We suggest a solution to this problem by introducing an adaptive segmentation approach which allows us to identify customer neighbourhood-based segments using the *similarity measure*. This measure is defined over a predictive variable space. The set of predictive variables is determined during a cross-validation procedure using rank correlations between the observed and predicted revenues as optimisation criteria. A one-period-ahead revenue is forecasted for each customer using a predictive probability distribution based on customers exhibiting similar behavioural characteristics, with bootstrap simulations being employed for a multi-period forecast. The model is developed and implemented for a UK bank.

* This research is supported by the Knowledge Transfer Partnership between the University of Leeds and National Australia Group Europe, trading as Clydesdale Bank and Yorkshire Bank. The authors thank Martin Allton, Head of Customer Knowledge, and Lucy Marshall, Customer Insight Manager, of National Australia Group Europe for their valuable contribution to this project.

1 Introduction

Customer Lifetime Value (CLV) is a well established concept in academic and business literature indicating future customer profitability for an organisation. Organisations aim at maximising customer value over the period of his or her relationship with a company. To this end, CLV offers a unique decision support tool which gives “the same focus throughout the different decision making areas of the organisation” (Thomas, 2000). Applications of this concept include customer relationship management (CRM) which seeks to improve the long-term retention and profitability of both currently highly profitable customers and also of less profitable or even unprofitable customers with a capacity to increase their profitability in the future (Zeithaml, Rust, and Lemon, 2001). Measuring customer profit generating potential along with credit scoring, which has been long used as a decision support tool in itself, is increasingly becoming an integral part of the companies’ lending decisions focusing on a true company’s objective of profit maximisation (Oliver, 1993, Fishelson-Holstine, 1998, Thomas, 2000, and Finlay, 2009). Another important application of customer lifetime value modelling is in customer segmentation where it facilitates the differentiation of levels and volume of customer servicing according to the profitability levels and cost optimisation (Zeithaml, 2001). Implementation of our CLV modelling approach for the UK bank led to better understanding of needs and preferences of some customer groups and to the development of tailored customer propositions.[1]

Companies have wrestled with the concept of CLV. They have been facing a number of problems in the context of the multiservice financial organisation. Firstly, the multidimensional nature of customer behaviour introduced challenges for the behavioural models as not only future purchase decisions but also their volumes need to be predicted (Donkers, Verhoef and de Jong, 2007). Secondly, customers often purchase more than one product and these purchasing decisions are not independent; these inter-dependencies should be taken into account when modelling customer lifetime value (Kamakura, Ramaswami and Srivastava, 1991, Kamakura et al., 2003, Knott, Hayes and Neslin, 2002, and Li, Sun and Wilcox, 2005). Thirdly, retail financial services organisations offer a wide variety of financial products (some rather complex), which differ in their nature and also in their revenue generating pattern. Finally, at any point in time a customer can switch between products or even between product providers (Kamakura et al., 2003). Intense competition and technological advances enabled historically monogamous retail bank customers (i.e., those conducting business only with one provider), to increasingly move towards the

[1] A segment was identified in the over 50s customer age group with high future revenue generating potential which was subsequently recognised by the organisation as a prime target for product cross-sale and customer acquisition and retention activities. This understanding has also contributed to the organisation’s compliance with the regulatory requirement of treating customers fairly by the UK Financial Services Agency.

“always a share” relationship in which a customer simultaneously conducts business with several financial services providers and can relatively easily switch between them. Due to these specifics of the retail financial industry, only a limited number of existing modelling approaches can be applied in this context.

We propose a novel approach to the modelling of customer lifetime value which is based on adaptive customer segmentation using a “nearest-neighbour” concept. Our approach identifies customer neighbourhoods (local segments) which are (1) small enough to ensure customer homogeneity by capturing only customers with similar characteristics and past behaviour, and (2) sufficiently large to ensure robust forecasts of future behaviour. The segmentation uses a *similarity measure* defined over a predictive variable space. The optimal local segment size and set of predictive variables are determined during a cross-validation procedure using rank correlations between the observed and predicted profitability as optimisation criteria. A one-period-ahead profitability is forecasted for each customer using a predictive probability distribution estimated over the population from his/her local segment containing customers exhibiting similar behavioural characteristics. Multi-period forecasts are produced by convolution of one-period predictive probability distributions which is implemented using bootstrap simulations.

Our segmentation approach allows the re-aggregation of small customer segments into larger ones in a flexible way that can be driven by the business decisions for which the segments are required. It can also be applied to a range of predictive tasks, such as the forecasting of the conditional profitability (given some other customer characteristics, e.g., credit risk) or prediction of other customer-related characteristics (e.g., product purchasing decisions or account balances). This provides a powerful tool in the development of tailored customer acquisition and retention strategies. The ranking approach offers benefits during times of significant economic change.

Our model is implemented using customer data for a UK retail bank for the period 2005-2008 to predict and validate customer profitability. We are able to make robust predictions of individual customer revenues using a small number of variables.

The remainder of this paper is structured as follows. Section two offers a short review of literature leading to our model. The details of our modelling approach are provided in section three. Section four offers an implementation algorithm which enhances the computational efficiency of our approach. The estimation results and discussion of potential applications are given in sections five and six correspondingly. Section seven concludes.

2 Review of previous studies leading to our model

Jain and Signh (2002) and Gupta et al. (2006) provide an overview of the existing lifetime value models. Empirical evidence in the context of a multiservice financial company (Donkers et al., 2007) suggests that currently available complex service-level models of customer behaviour do not offer an advantage over simple models using aggregate customer data when predicting individual customer lifetime value. These simple models are based on a strong limiting assumption that customer profitability is constant over time or can be represented by a linear function of past profitability (e.g., Berger and Nasr, 1998, and Malthouse and Blattberg, 2003). This assumption can be satisfactory in the context of retail banking only for customers who continuously hold one or more long-term products (e.g., a savings account and a personal loan or a mortgage) over a number of years, so that the company's future profits are determined by regular (constant, increasing or decreasing at a certain rate) payments and a constant profitability margin. Both assumptions are not valid when a customer purchases a new product, or switches products or product providers before product maturity, or when interest rates (and hence, sales margins) change. From a business perspective one is most interested in customers with a high propensity to buy new products or to attrite, as immediate action is required in these cases to ensure the maximisation of their CLV.

More complex models of customer behaviour which are used for the customer lifetime value predictions adapt the univariate or multivariate probit modelling frameworks to predict next period purchase probabilities. The multivariate probit model accounts for dependencies between the purchasing decisions arising either from a hierarchy in the decisions to add new services to the current portfolio, or from cross-sales promotions or sale of financial packages (e.g., Kamakura et al., 1991 and 2003). The main weaknesses of this model class are that (1) only one-period ahead prediction can be made and (2) only a purchase decision is predicted and a separate model is needed to estimate the purchase value, which varies greatly in the context of a retail bank and introduces additional complexity to the model. This complexity contributes to larger prediction errors and a weaker forecasting ability for these models of individual customer profitability compared to simpler models (Donkers et al., 2007). Another widely discussed approach used in customer relationship and behavioural models employs the Markov Chain methodology (e.g., Morrison et al., 1982, and Pfeifer and Carraway, 2000). This class of models allows for additional flexibility in modelling customer behaviour when compared to earlier models. An important restrictive limitation of the Markov Chain methodology, however, is that the robustness of profitability estimation depends on the existence of a relatively small number of meaningful and homogenous customer segments. The latter condition is difficult to satisfy in the retail banking industry given that there is a significant diversity of customer purchasing behaviour, which may result in a wide range of revenues, even for customers who are similar in terms of characteristics such as age, tenure as a customer, earning capacity etc. which are often used as a basis for customer segmentation. As a result, the robustness of the estimation suffers. Also, allocating customers to a limited number of segments may lose a lot of information, which again affects the robust estimation of model parameters, which may therefore not be enough to satisfy business needs in terms of differentiating between groups of customers for marketing decisions.

3. Modeling customer behaviour

In this section we describe our approach to predicting the customer lifetime value of individual customers in the context of a multiservice financial organization. As discussed above, in this industry a customer's relationship with a company is typically long-term and customer purchasing behaviour is rather complex. Also, due to the specifics of retail banking, a purchase of the same product (say, a personal loan or fixed-term deposit) by two different customers may result in substantially different revenues as the revenues vary not only with the decision to purchase but also across the wide range of possible purchase volume. Existing models of customer behaviour using a Markov Chain methodology and probabilistic models developed within a probit modeling framework do not offer a satisfactory solution to this problem. We propose an adaptive segmentation approach which uses information from the historic probability distributions of a set of predictive variables for a close customer neighborhood to predict the future revenues associated with a given customer. Our approach allows for a wide diversity of customer purchasing behaviour without compromising the robustness of the estimation. Our approach is explained in detail next.

3.1 General CLV model

We start with a general definition of customer lifetime value as the net present value of future cash flows associated with a customer over the time period of her relationship with a company or over a given time horizon. Many articles provide slightly differing formulations of the classic customer lifetime value model (Berger and Nasr, 1998, Jain and Singh, 2002, Reinartz and Kumar, 2003, and Gupta et al., 2006, among others) which can be summarised as follows.

$$CLV^{i} = \sum_{\tau=0}^{T} (R_{\tau}^{i} - C_{\tau}^{i})(1 - q_{\tau}^{i})\, D_{\tau} - AC_{0}^{i}\,, \qquad (1)$$

where R_τ^i denotes the predicted revenue from customer i in period τ given that the customer continues his or her relationship with the company in period τ and C_τ^i is the direct costs of servicing the customer in this period. AC_0^i denotes a cost of acquisition for a new customer. D_τ is the discount factor in period τ and T is the number of periods for which CLV is estimated. q_τ^i is the projected probability that customer i may terminate the relationship with a company in period τ, $\tau < T$; this is sometimes called the customer attrition rate.

In the current application, which is presented in this paper, we focus on predicting future revenues from a customer R_τ^i assuming that direct costs C_τ^i associated with servicing a customer during $0 \le \tau \le T$ and costs of customer acquisition AC_0^i are known. We address the estimation of the customer attrition rate q_τ^i in the next section. Considering that our approach can be applied to a wide range of forecasting problems, we present a general formulation for the proposed approach next.

3.2 Adaptive segmentation approach

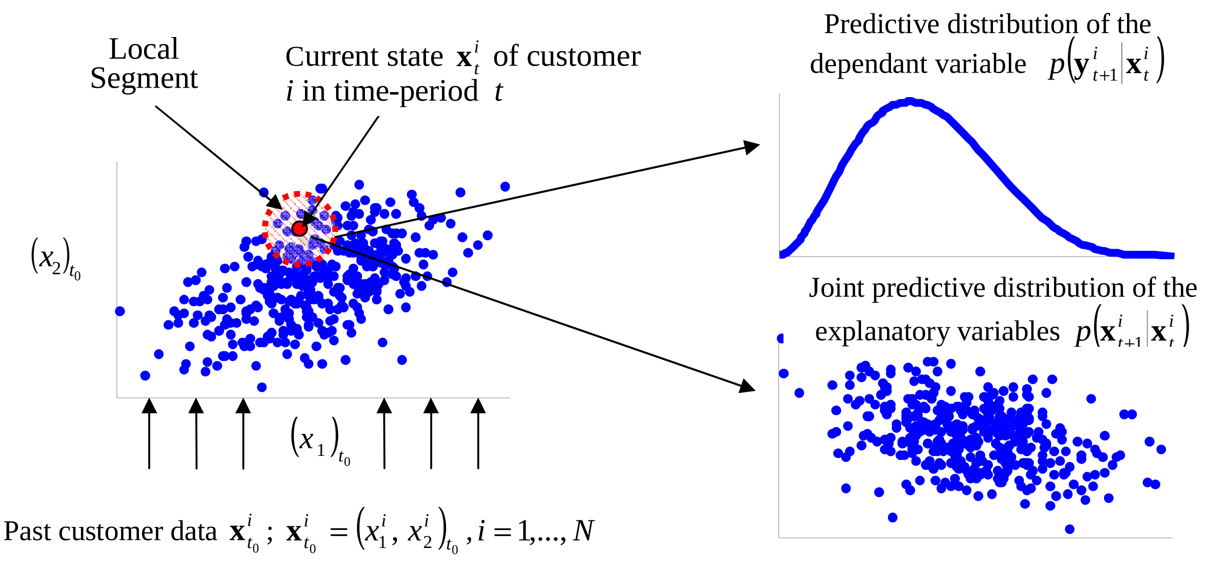
For each customer, we aim to obtain the conditional probability distribution of observing the vector of dependent variables $\mathbf{y}_{t+\tau} = (y_1, \ldots, y_M)_{t+\tau}$, $\tau = 0, 1, 2, \ldots, T$. $\mathbf{y}_{t+\tau}$ can be generally a multidimensional vector and include both continuous and cardinal variables. In the current application it consists of the customer revenue-variable $R_{t+\tau}^i$. $\mathbf{x}_t$ is a vector of K predictive variables, $\mathbf{x}_t = (x_1, \ldots, x_K)_t$ which can include any elements of the historic information set for a customer available at time *t*. $\mathbf{x}_t$ provides the explanatory variables in our model. In the current application, it contains a set of variables from the company's customer database which are predictive of customers' future revenue (and of other, if any, dependent variables in $\mathbf{y}_{t+\tau}$ for the general case) and may include variables observed at all or some of time periods $t, t-1, t-2, \ldots, t-L$. t is the most recent period and L is the number of periods for which historic customer data is available. Typical predictive variables, for example, might include current and past customer revenues, his or her age, tenure with the company, credit score, and current and past product holdings.

The conditional probability distribution of observing vector of dependent variables $\mathbf{y}_{t+\tau}$ given the current customer state $\mathbf{x}_t$ is $p(\mathbf{y}_{t+1}|\mathbf{x}_t)$, $p(\mathbf{y}_{t+2}|\mathbf{x}_t)$, …, $p(\mathbf{y}_{t+T}|\mathbf{x}_t)$. We call these distributions *predictive distributions* from now on. Note that the probability $p(\mathbf{y}_{t+\tau}|\mathbf{x}_t)$ is conditional on current customer state $\mathbf{x}_t$

and also on a customer continuing his or her relationship with the company until and throughout period $t+\tau$.

Because the analytical form of the predictive probability distributions are not known, it is suggested to estimate these distributions empirically using an *adaptive segmentation approach*. Given a currently available information $\mathbf{x}_t$ for a customer, we locate a group of customers that were in a similar state (and had a similar history, if historic variable observations are included in $\mathbf{x}_t$) one or more periods ago; we name this time-period as the basis period t_0. This group is called here *a local segment* or *customer neighborhood*. The empirical conditional distribution of the dependent variable for the *local segment* at t_0+1 serves as an estimate of the one-period-ahead predictive distribution of the dependent variable. The empirical joint distribution of the explanatory variables within a local segment is used as an estimate of the one-period-ahead predictive distribution of the explanatory variables which is used for multi-period predictions. Figure 1 illustrates this concept.

Figure 1 Estimation of the empirical predictive distributions with a bivariate vector of explanatory variables (K=2). *N* is the number of customers in the customer database



In this approach customers with similar characteristics and past behaviour are expected to exhibit similar behaviour in the future. The size of the local segment is chosen so that it is (1) small enough to ensure customer homogeneity within the local segment by capturing only customers with similar characteristics and past behaviour, and (2) sufficiently large to ensure robust forecasts of future behaviour. The local segmentation uses a *similarity measure* defined over the predictive variable space to locate the customer neighbourhood.

The *similarity measure D* is defined as the weighted Euclidean distance between the predictive variables of two customers for which information is available at time *t* and which are contained in $\mathbf{x}_t$. This measure is used to derive the neighbourhood of a customer:

$$D\left(\mathbf{x}_t^i, \mathbf{x}_{t_0}^j\right) = \sqrt{\sum_{k=1}^{K} \left[w_k \left(x_{k,t}^i - x_{k,t_0}^j\right)\right]^2} \qquad (3)$$

Weights w_k, $k = 1,...,K$, are used to normalize the explanatory variables which all have different measurement scales. For example, customer age may vary between 18 and 100 years whereas the spot balance on a money transmission account can be up to several million pounds. Without weighting, the *similarity measure* will be dominated by the variable with the largest measurement scale.

For continuous dependent variables, e.g., the period revenue, the empirical predictive distributions are used to estimate mean and median values and confidence intervals. For cardinal dependent variables, a frequency of observing a customer in the local segment with a given value of the dependent variable at $t_0 + 1$ is employed as an estimate of the corresponding probability. This estimate is then compared against some threshold to provide a value forecast. The threshold is estimated during the model validation stage using a ROC-analysis framework. ROC[2] analysis is a standard tool originating from signal detection theory. It is used across a wide range of practical applications for organizing and selecting classifiers on the basis of their performance (Fawcett, 2006).

3.3 Multi-period-ahead forecast

Under the assumption of time-homogeneity, a multi-period-ahead predictive distribution can be obtained from the one-period-ahead predictive distributions for dependent and explanatory variables. For example, the two-period-ahead predictive marginal probability distribution can be obtained as a convolution of the corresponding one-period-ahead distributions for dependent and explanatory variables:

$$p\left(\mathbf{y}_{t+2} \middle| \mathbf{x}_t\right) = \int p\left(\mathbf{y}_{t+2}, \mathbf{x}_{t+1} \middle| \mathbf{x}_t\right) d\mathbf{x}_{t+1} = \int p\left(\mathbf{y}_{t+2} \middle| \mathbf{x}_{t+1}\right) p\left(\mathbf{x}_{t+1} \middle| \mathbf{x}_t\right) d\mathbf{x}_{t+1}$$

Following a similar logic, the multi-period-ahead predictive distribution is obtained by the convolution of the corresponding one-period-ahead distributions (e.g., Andersen, Bollerslev, Christoffersen and Diebold, 2006, p.810):

[2] ROC stands for Receiver Operating Characteristics and it is also known as Relative Operating Characteristics

$$p(\mathbf{y}_{t+\tau}|\mathbf{x}_t) = \iint \ldots \int p(\mathbf{y}_{t+\tau}|\mathbf{x}_{t+\tau-1}) p(\mathbf{x}_{t+\tau-1}|\mathbf{x}_{t+\tau-2}) \ldots p(\mathbf{x}_{t+1}|\mathbf{x}_t) d\mathbf{x}_{t+\tau-1} \ldots d\mathbf{x}_{t+1} \qquad (2)$$

It follows from (2) that knowing one-period-ahead predictive distributions $p(\mathbf{y}_{t+\tau}|\mathbf{x}_{t+\tau-1})$, $0 \le \tau \le T$, for a dependent variable and $p(\mathbf{x}_{t+\tau}|\mathbf{x}_{t+\tau-1})$ for explanatory variables is sufficient to obtain multi-period-ahead predictions.

As in case of one-period-ahead forecast, empirical predictive probability distributions are employed for multi-period ahead predictions. Our empirical approach is first illustrated for a two-period-ahead forecast. Given a current customer state $\mathbf{x}_t$, a random vector $\mathbf{x}_{t+1}$ is simulated from the one-period predictive distribution $p(\mathbf{x}_{t+1}|\mathbf{x}_t)$. This can be done either by using re-sampling techniques such as bootstrap, which is employed in this paper, or by sampling from an approximating analytical distribution, e.g., by using copulas or kernel smoothing. Using the simulated values from $\mathbf{x}_{t+1}$, we simulate the corresponding values of $\mathbf{y}_{t+2}$ for the next period using the analogous technique as in the previous step. The resulting pair $\mathbf{x}_{t+1}, \mathbf{y}_{t+2}$ has joint probability distribution $p(\mathbf{y}_{t+2}|\mathbf{x}_{t+1}) p(\mathbf{x}_{t+1}|\mathbf{x}_t)$. The marginalization over $\mathbf{x}_{t+1}$ is achieved by pooling $\mathbf{y}_{t+2}$ for all simulated values of $\mathbf{x}_{t+1}$. A similar approach is used to estimate the empirical predictive distribution for periods $t+\tau$, $0 \le \tau \le T$: given simulated values for $\mathbf{x}_{t+\tau-1}$, a random vector $\mathbf{y}_{t+\tau}$ is obtained from the one-period predictive distribution $p(\mathbf{y}_{t+\tau}|\mathbf{x}_{t+\tau-1})$.

3.3 Real-life considerations for assigning variable weights in the similarity measure

Errors and inaccuracies are unavoidable in real-life company data bases due to the imperfections in data collection processes (SAS reference). Thorough data cleaning might prove too expensive and impractical for many organizations. Our method for assigning weights w_k in (3), therefore, is designed using practical considerations and aims to minimise the impact of these errors and inaccuracies on the predictive power of our model. Our analysis of the company data showed many data errors represent outliers which may increase the variable range to up to 5%. The 95% variable range, therefore, is representive of the 'clean' variable range for most continuous variables. This variable range is used in our weighting and included in the denominator in (4). Weights defined in this way penalize outliers whose weighted difference entering Equation (3) will be more than 1.

The weights for continuous variables in (3) are defined so that the variable values in the 'clean' range are constrained to the interval of [0, 1].

$$w_k = \frac{1}{Percentile\left(x_{k,t_0}, 0.975\right) - Percentile\left(x_{k,t_0}, 0.025\right)} \qquad (4)$$

For cardinal explanatory variables, $w_k = 0$ for equal values and $w_k = \infty$ for non-equal values so that only customers with a given value of a cardinal variable enter a local segment.

3.4 Discussion of adaptive segmentation approach

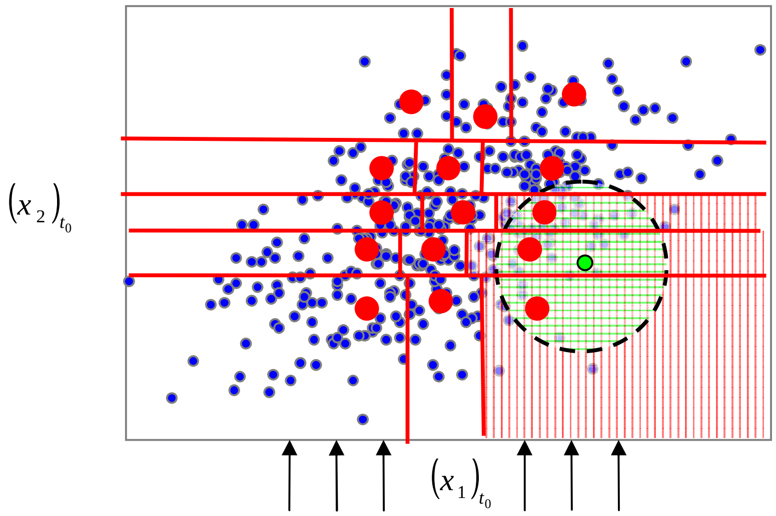
Our modeling approach with adaptive segmentation offers a number of advantages over the existing approaches in the context of the multiservice financial industry. First, our approach works with various shapes of variable distribution and also with different correlation structures (no limiting assumptions are imposed). This is important in our application given that most variables are not normally distributed and exhibit a non-trivial dependence structure (e.g., strong non-linear correlation and autoregressive effects are detected in our dataset). Second, full information, contained in the variable probability distribution, is preserved which allows the estimation of confidence intervals and other important statistics. Third, unlike Markov Chain and other probabilistic models using customer segments, our model adaptively seeks for homogenous customer segments without loss of information about variables distribution. Fourth, the model can work with partial information and missing variables still producing a meaningful forecast if some values for predictive variables are not available. This allows the prediction of future revenue for new-to-bank customers, for whom only partial information is initially available, and also overcomes the problem of missing values which is relevant to many companies due to the imperfections of the company data collection process. Finally, the effect of errors and inaccuracies present in the raw company data is minimized by scaling in our adaptive segmentation approach.

On the downside, given that our model is non-parametric an additional analysis is required to draw generalizing inference about the relationships in the data. Also, our adaptive segmentation approach is computationally intensive. The next section deals with the latter drawback.

4 Optimisation of computational efficiency

Our adaptive segmentation approach can be computationally intensive. The following algorithm for prior coarse segmentation substantially reduces the computational load. Prior to locating a customer's *local segment*, we suggest applying a coarse segmentation of the customer dataset for *local segmentation* using the set of explanatory variables. The centres and boundaries of coarse segments need to be recorded at this stage. The computational load is substantially reduced because the *similarity measure* is calculated only for the central points of the coarse segments instead of all entries in the dataset. At the next stage, only coarse segments which are the closest to the current state of the customer in question are selected; entries from these segments are pooled together and a local segment is chosen out of the pooled entries. Figure 2 illustrates this algorithm.

Figure 2
Adaptive segmentation: optimisation of computation
Red dots and red lines show coarse segments' centres and boundaries; the green dot corresponds to the current state of a customer whose revenue is forecasted; the red area covers the closest coarse segments and green area is the local segment



Past customer data $\mathbf{x}^i_{t_0}$; $\mathbf{x}^i_{t_0} = \left(x^i_1, x^i_2\right)_{t_0}$, $i = 1,\ldots, N$ available in time-period t_0

The proposed approach can potentially exclude a small number of entries on the external border of a local segment from the area available for local segmentation, although they belong to the local segment. Such excluded entries are illustrated by the top part of the green-shaded circle outside of the red-shaded area in Figure 2; this is the upper part of the local segment but is not included in the selection of coarse segments. The excluded area, however, is very small and it has a negligible effect on the estimated values.

The segment size is the same across all coarse segments to ensure equal representation. The coarse segment size is chosen with an objective to minimize the amount of computation at the segmentation stage. It is shown below that a segment size which is close to $\sqrt{N}$ satisfies this objective (here N is a number of customer entries in the local segmentation dataset in time period t_0). Denoting the number of coarse segments as S and the number of customers in each segment as s will give $N = s \cdot S$. Set $s \equiv \alpha\sqrt{N}$ to obtain $S = \dfrac{\sqrt{N}}{\alpha}$. The number of computations of the *similarity measure* is:

$$S + s = \sqrt{N}\left(\alpha + \frac{1}{\alpha}\right) \qquad (5)$$

The number of comparisons which are needed to sort distances to the coarse segment centers and also to sort customers within the coarse segments would be $S\log_2 S + s\log_2 s$ (Press et al., 1992). Substituting values of s and S will give:

$$S\log_2 S + s\log_2 s = \sqrt{N}\alpha\log_2\left(\sqrt{N}\alpha\right) + \frac{\sqrt{N}}{\alpha}\log_2\frac{\sqrt{N}}{\alpha} \qquad (6)$$

Differentiation of (5) and (6) in respect to α and setting the resulting expressions to zero gives the result that the minimal amount of computation is achieved with $\alpha = 1$.

5 Estimation results

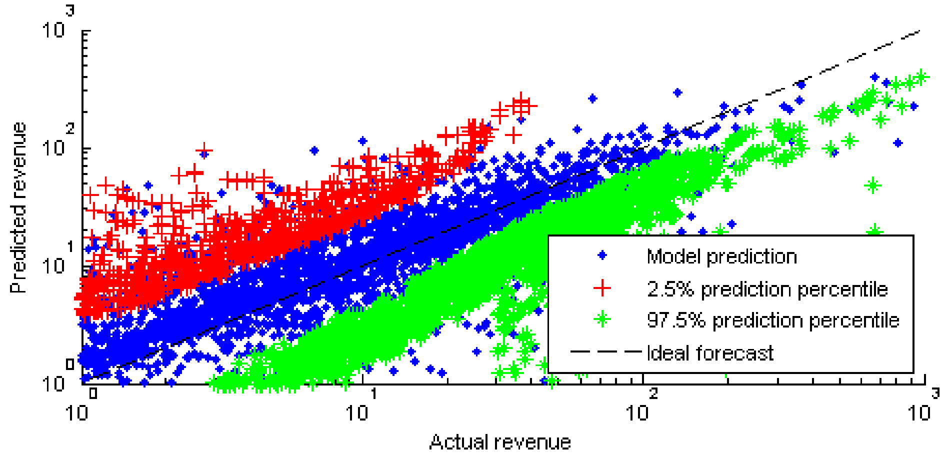
This section presents the model estimation and validation results. Model validation in our approach has two objectives: (1) to choose a set of predictive variables with the best predictive power and (2) to evaluate the predictive performance of the model. The model estimation, prediction and validation are performed using historic revenues and other retail customer data for 467,789 retail banking customers of National Australia Bank Europe over the period 2005-2008. In the process of validation the model prediction for one-period ahead and two-periods ahead has been assessed against actual observed values. We have been limited by data availability from performing the validation of predictions for a higher number of periods ahead.

The main purpose of this study was to predict the customer revenue as an outcome of the customer behaviour. The customer revenue, expressed in absolute figures, however, is affected by external factors even if customer behaviour stays the same in different time periods. For example, the change in the base rate of the Bank of England has a direct impact on the sales margin of the financial services provider. Firstly, many long-term financial products are sold with a fixed interest rate and products sold prior to the change in the base rate will bear the old price (either increasing or decreasing the amount of sales margin depending on the direction of this change). Secondly, interest rates on retail financial products do not always adjust in a direct correspondence with changes in the base rate. The latest significant mismatch between the Bank of England rate, costs of wholesale borrowing and interest rates on retail financial products was observed throughout the second half of 2008 during the peak of the financial crisis. Intense competition may also cause a financial provider to reduce sales margins.

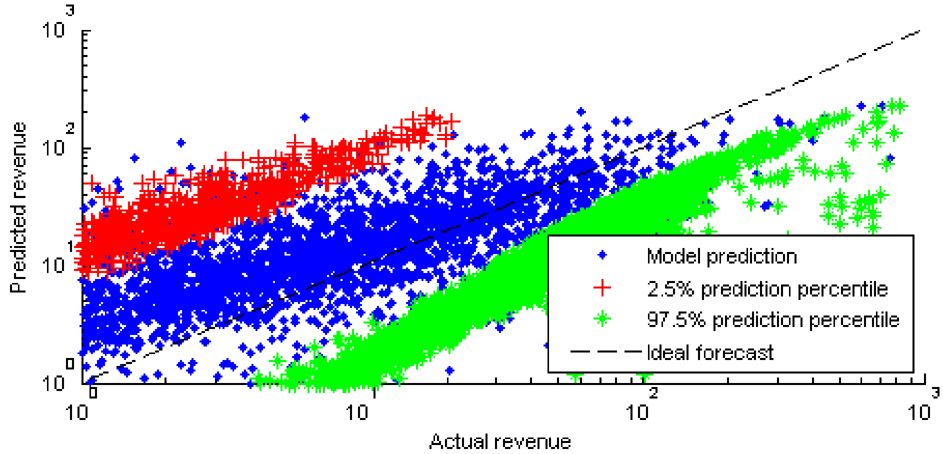
To account for changes in future revenue due to external factors, the set of predictive variables in our model is chosen during the process of validation by maximizing the rank correlation statistics between the predicted and actual values. Kendall’s tau has been used for this purpose. This statistic has an advantage of being insensitive to any variable transformation which does not change the ordering in the population.

Using this statistic in choosing the model ensures that the final model best preserves the relative ordering in the dependent variable (the revenue from a customer in our case) and accurately predicts customers with higher future revenue versus customers with less future revenue, even if the revenue value is affected by external factors in future.

Figure 3. Actual versus predicted customer revenue; logarithmic scale
One-period-ahead prediction



Two-period-ahead prediction



Our model gives robust predictions of customer revenue for one and two periods ahead. In Figure 3, the slope of the line around which the model predictions are grouped indicates the predictive power of our model.[3] Predictions along the diagonal line indicate a perfect forecast when predicted revenue is exactly equal to actual revenue. Low predictive power results in a near-zero slope when predicted revenues are scattered around the mean revenue in the population. A higher predicted than actual revenue value for customers in the lower revenue range could be due to the decrease in sales margins on many retail

[3] We cannot reveal the actual revenue figures and also the predictive variables in our model due to the commercial sensitivity of this information.

products which happened in the second half of 2007 and 2008. This trend should not affect the robustness of prediction of the relative revenue ranking.

Table 1
Model statistics: validation of one and two-periods-ahead predictions of the revenue from individual customers using actual values of revenue in 2007 and 2008 correspondingly

Periods ahead	Kendall's τ	Kendall's Z	Concordant pairs, Percent	Discordant pairs, Percent	Gamma	Somers D(R|C)	Somers D(C|R)	Mean error	Median error	St. Dev.
One Period	0.71	74.90	0.83	0.13	0.72	0.69	0.72	0.29	0.10	27.34
Two periods	0.62	66.06	0.78	0.17	0.64	0.61	0.64	2.59	1.89	35.01

The rank correlation (Kendall's tau) is significant for both predicted periods and higher for a one-period-ahead prediction (0.71) than for two-periods-ahead prediction (0.62). Median prediction error is also greater for the two-period-ahead prediction. This is expected as more distant periods are more difficult to predict. Also, our data for this validation comes from 2008 which coincides with the banking crisis which saw a market-wide misalignment between the cost of funds and the lending interest rate and also unusually high growth in bad and doubtful debts. Our model demonstrates a strong performance in this non-trivial environment.

We employed Kendall's tau, indicating rank correlation between the actual and predicted values, to assess the sensitivity of the accuracy of prediction to the size of the local segment. Kendall's tau, at first, increased with an increase in the local segment size, then reached a maximum value at around 1000, followed a plateau at the maximal level and eventually started decreasing with further increases in the local segment size. This pattern can be explained as follows. With a small segment size, the empirical predictive distribution is dominated by customers that closely match the current state and, thus, tend to exhibit a very similar behaviour in the next period. Over-expanding the size of the local segment, on the other hand, allows too many customers with different behavioural characteristics into a local segment shifting the predictive distribution for this segment towards the population mean. Both situations lead to an inferior predictive power of the model. Based on these considerations, a local size segment has been set at 1000 customers.

6 Other applications

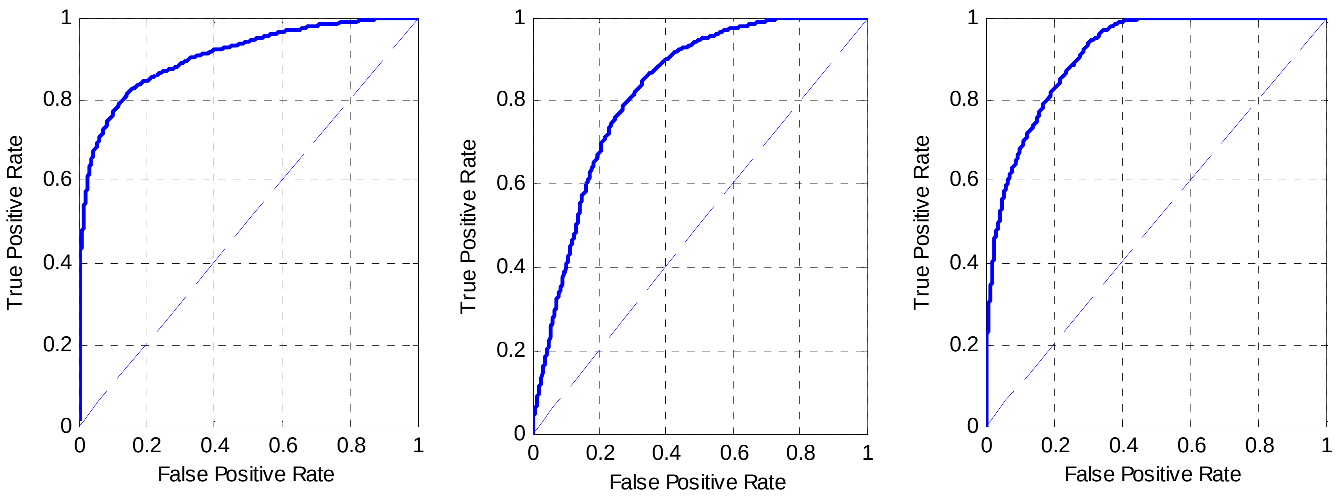
With a view to directing customer management effort and also marketing spend towards the right customers, the company management is interested in identifying customers whose revenue in likely to increase or decrease significantly in future periods. We therefore can employ our modeling approach to another application: prediction of binomial variables indicating whether the revenue from a customer is likely to increase (decrease) above (below) a given threshold or remain within the threshold in the next period(s) (Table 2 and Figure 3). Given that companies often do not possess information on the details of their customers' activities with other providers, this objective can be rather challenging when traditional models are used. By matching a customer to the right 'local segment', our adaptive segmentation approach allows the extraction of 'additional information' to the directly-measurable information, that companies normally possess, from the pattern in behaviour of the customers in the local segment possessing similar characteristics and past behaviour.

Table 2
Model statistics: validation of one-period ahead forecast of binomial variables indicating (1) an increase in revenue above the minimum threshold ('jump up'), (2) decrease in revenue below the minimum threshold ('jump down') and (3) change in revenue within the threshold ('stable')

Forecasted variable	Area Under Curve	Accuracy
Stable	0.9068	0.8488
Jump Up	0.8256	0.7566
Jump Down	0.9100	0.8690

Figure 3
ROC curves, representing Relative Operating Characteristics, for the one-period-ahead prediction of customer behaviour

'Stable': change in revenue within the threshold | **'Jump up'**: increase in revenue above the minimum threshold | **'Jump down'**: decrease in revenue below the minimum threshold



Similarly, our approach can be employed for prediction of customer attrition rates which act as an input variable for CLV estimation in (1). Another potential application for our adaptive segmentation approach would be in predicting the propensity for a customer to purchase a financial product and also a purchase volume.

In general terms this modelling approach can be applied in a range of predictive applications, whether of values or states, and its ranking approach makes it useful in situations where environmental characteristics are unstable or unpredictable.

7 Conclusion

This paper proposes an adaptive segmentation approach to the modelling of lifetime value for individual customers in a multiservice financial organisation. The main contribution of our modelling approach is that unlike Markov Chain and other probabilistic models using customer segments, our model adaptively seeks to locate homogenous customer segments without loss of information about variable distributions compromising the accuracy of prediction. This is important because customer behaviour is very varied in this sector even for customers who are similar in terms of the characteristics which are normally used for customer segmentation (e.g., customer age product holdings and personal income). Customers purchase more than one product or service, and these purchases are not independent. The revenue from a customer is determined not only by his or her purchase decision but also by the purchase amount which varies widely for different customers.

Other advantages of our modelling approach are that it does not require assumptions about the shape of the variable distributions or the correlation structure between them; the model can work with partial information and missing variables, producing a meaningful forecast if some values for predictive variables are not available. This allows the prediction of future revenue for new-to-bank customers, for whom only partial information is initially available, and also overcomes the problem of missing values which is relevant to many companies due to the imperfections of the company data collection process.

We applied our model to the estimation and validation of future customer revenue in the UK retail bank over 2005-2008. Our model gave robust predictions of the revenues from individual customers and also of significant changes in their revenues using a small number of predictive variables. Validation of our 2-period-ahead forecast using 2008 data, which coincided with the crisis year in banking, has confirmed that our relative ranking approach offers advantage during times of significant economic change. Other potential applications include the prediction of other customer-related characteristics which could be either continuous or cardinal variables. Our approach allows the re-aggregation of small customer segments into larger ones in a flexible way that can be driven by the business decisions for which the segments are required. This provides a powerful tool in the development of tailored customer acquisition and retention strategies.

References

Andersen, T.G., Bollerslev, T., Christoffersen, P.F., Diebold, F.X., 2006. Volatility Forecasting, In: Elliott, G., Granger, C.W.J., Timmermann, A. (Editors), Handbook of Economic Forecasting, Volume 1, North Holland, Amsterdam, p.810.

Berger, P.D., Nasr, N.I., 1998. Customer Lifetime Value: Marketing Models and Applications. Journal of Interactive Marketing 12, 17-29.

Donkers, B., Verhoef, P.C., de Jong, M., 2007. Modeling CLV: A Test of Competing Models in the Insurance Industry. Quantitative Marketing and Economics 5, 163-190.

Fawcett, T., 2006. An Introduction to ROC Analysis. Pattern Recognition Letters 27, 861-874.

Finlay, S., 2009. Credit scoring for profitability objectives. European Journal of Operational Research. Forthcoming.

Fishelson-Holstine, H., 1998. Case studies in credit risk model development. In: Mays, E. (Editor), Credit Risk Modeling, Glenlake Publishing, Chicago, pp. 169–180.

Gupta, S., Hassens, D., Hardie, B., Kahn, W., Kumar, V., Lin, N., Ravishanker, N., 2006. Modeling Customer Lifetime Value. Journal of Service Research 9, 139-155.

Jain, D., Singh, S.S., 2002. Customer Lifetime Value Research In Marketing: A Review And Future Directions. Journal of Interactive Marketing 16, 34- 46.

Kamakura, W.A., Ramaswami, 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., de Rosa, F., Mazzon, J.A., 2003. Cross-Selling Through Database Marketing: A Mixed Data Factor Analyser for Data Augmentation and Prediction. International Journal of Research in Marketing 20, 45-65.

Knott, A., Hayes, A., Scott, A.N., 2002. Next-Product-to-Buy Models for Cross-Selling Applications. Journal of Interactive Marketing 16, 59-75.

Li, S., Sun, B., Wilcox, R.T., 2005. Cross-Selling Sequentially Ordered Products: An Application to Consumer Banking Services. Journal of Marketing Research XLII, 233-239.

Malthouse, E.C., Blattberg, R.C., 2003. Evaluating the Predictive Accuracy of Customer Long-Term Values for Customer Relationship Management Strategies. *Working Paper*, North Western University.

Morrison, D.G., Chen, R.D.H, Karpis, S.L., Britney, K.E.A, 1982. Modelling Retail Customer Behaviour at Merrill Lynch. Marketing Science 1, 123-141.

Oliver, R.M., 1993. Effects of calibration and discrimination on profitability scoring. In: Proceedings of Credit Scoring and Credit Control III, Credit Research Centre, University of Edinburgh

Pfeifer, P.E., Carraway, R.L., 2000. Modeling Customer Relationships as Markov Chains. Journal of Interactive Marketing 14, 43-55.

Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P., 1992. Numerical Recipes in C: The Art of Scientific Programming, 2nd edition. Cambridge University Press.

Reinartz, W., Kumar, V, 2003. The Impact of Customer Relationship Characteristics On Profitable Lifetime Duration. Journal of Marketing 67, 77-99.

Thomas, L.C., 2000. A survey of credit and behavioural scoring: forecasting financial risk of lending consumers. International Journal of Forecasting 16, 149-172.

Zeithaml, V., Rust, R.T., Lemon, K.N., 2001. The Customer Pyramid: Creating and Serving Profitable Customers. California Management Review 43, 118-142.