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

A. Audzeyeva, *University of Leeds & National Australia Group Europe, UK*
B. Summers, *University of Leeds, UK*
K.R. Schenk-Hoppé, *University of Leeds, UK*

CREDIT SCORING & CREDIT CONTROL XI CONFERENCE

Edinburgh, August 2009

Motivation & contribution

Motivation

Customer Lifetime Value: Key business tool for customer management

- Gives "the same focus throughout the different decision making areas of the organisation" (Thomas, 2000)
  - CRM: customer relationship management (Zeithaml et al., 2001)
  - Integral part of lending decisions (Oliver, 1993, Fishelson-Holstine, 1998, Finlay, 2009)
  - Customer segmentation for services differentiation & cost optimisation (Zeithaml et al., 2001)
- Challenges in CLV estimation in the multiservice financial industry:
  - Multidimensional nature of customer behaviour (Donkers et al., 2007)
  - Multiple non-independent purchases (Kamakura et al., 1991, 2003; Li et al., 2005))
  - Customers can easily switch between products & product providers

Contribution

- Propose a novel approach to CLV prediction based on adaptive segmentation
- Neighbourhood-based segments capture homogenous customers with similar characteristics & past behaviour
- Approach can be applied to a range of predictive tasks (e.g., product purchasing decisions and purchase volumes)
- Approach implemented for a UK retail bank

Literature

Empirical evidence for multiservice financial industry: current complex service-level models of customer behaviour do not offer advantage over simple models using aggregate data (Donkers et al., 2007)

- Simple models: customer profitability is constant over time or a linear function of past profitability (Berger and Nasr, 1998, and Malthouse and Blattberg, 2003)
- Complex models:
  - Multivariate probit (Kamakura et al., 1991 and 2003)
    - only 1-period prediction can be made
    - purchase volumes need to be predicted separately
  - Markov Chain Methodology (Morrison et al., 1982, and Pfeifer and Carraway, 2000)
    - depends on existence of a small number of meaningful segments

τ

General CLV model

$$CLV^i = \sum_{\tau=0}^{T} (R_\tau^i - C_\tau^i)(1 - q_\tau^i) D_\tau - AC_0^i$$

R_τ^i - predicted revenue from customer *i* in period τ given he/she continues relationship with the company in this period

C_τ^i - direct cost of servicing customer in period τ

AC_0^i - cost of acquisition of a new customer

D_τ - discount factor in period τ

q_τ^i - projected probability that customer *i* may terminate his/her relationship with a company in period τ , $\tau < T$

(Berger and Nasr, 1998, Jain and Singh, 2002, Reinartz and Kumar, 2003, & Gupta et al., 2006, among others)

Adaptive segmentation approach

Aim to obtain the conditional probability distribution

$$\mathbf{y}_{t+\tau} = \left(y_1, \ldots, y_M\right)_{t+\tau} \quad , \quad \tau = 0, 1, 2, \ldots, T$$

$\mathbf{x}_t$ is a vector of predictive variables: $\mathbf{x}_t = \left(x_1, \ldots, x_K\right)_t$

It can include any elements of the historic information set for a customer available at time *t*

The conditional probability distributions of observing the vector of dependent variables given the current customer state $\mathbf{x}_t$:

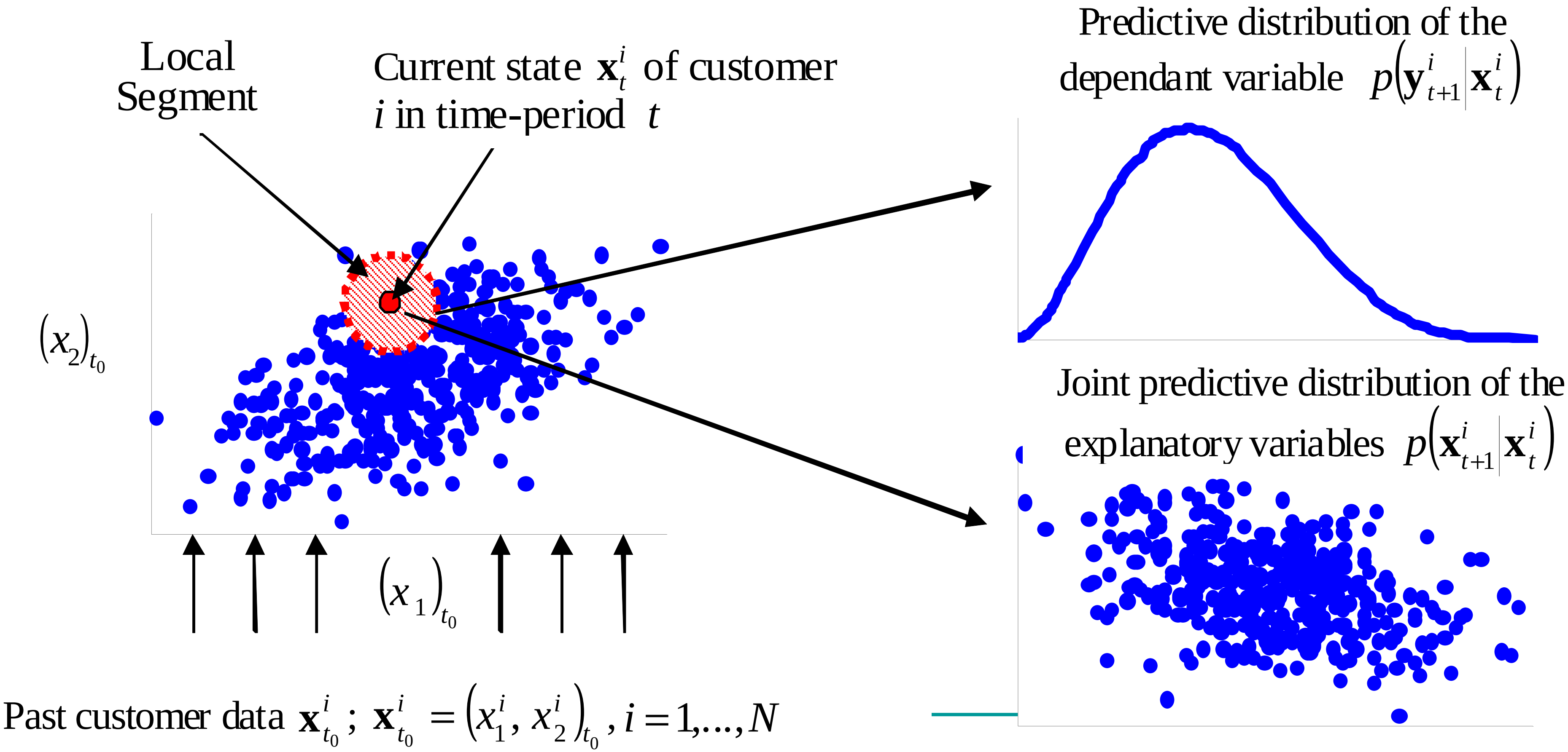
$$p\left(\mathbf{y}_{t+1} \middle| \mathbf{x}_t\right) \ , \ p\left(\mathbf{y}_{t+2} \middle| \mathbf{x}_t\right) \ , \ \ldots, \ p\left(\mathbf{y}_{t+T} \middle| \mathbf{x}_t\right)$$ - *predictive distributions*

Because the analytical form of *the predictive distributions* is not known, we estimate them empirically using *an adaptive segmentation approach*

Assumption: Customers with similar characteristics and past behaviour expected to exhibit similar behaviour in the future

Adaptive segmentation approach

Figure1. Estimation of the empirical predictive distributions with 2 predictive variables



Adaptive segmentation approach

- Size of the local segment:
  - small enough to ensure customer homogeneity within the local segment
  - sufficiently large to ensure robust forecasts of future behaviour
- Local segmentation uses a *similarity measure:*

$$D\left(\mathbf{x}_t^i, \mathbf{x}_{t_0}^j\right) = \sqrt{\sum_{k=1}^{K}\left[w_k\left(x_{k,t}^i - x_{k,t_0}^j\right)\right]^2} \quad w_k = \frac{1}{Percentile\left(x_{k,t_0}, 0.975\right) - Percentile\left(x_{k,t_0}, 0.025\right)}$$

Weights w_k are used to normalize the explanatory variables with different measurement scales

- Continuous dependent variables: the empirical predictive distributions used to estimate mean/ median values and confidence intervals
- Cardinal dependent variables: a frequency of observing a customer with a given value gives as an estimate of the corresponding probability. The threshold for a value forecast estimated during the model validation using a ROC-analysis framework.

Multi-period forecast

- Two-period-ahead predictive distribution

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

- Tau-period-ahead ahead predictive distribution:

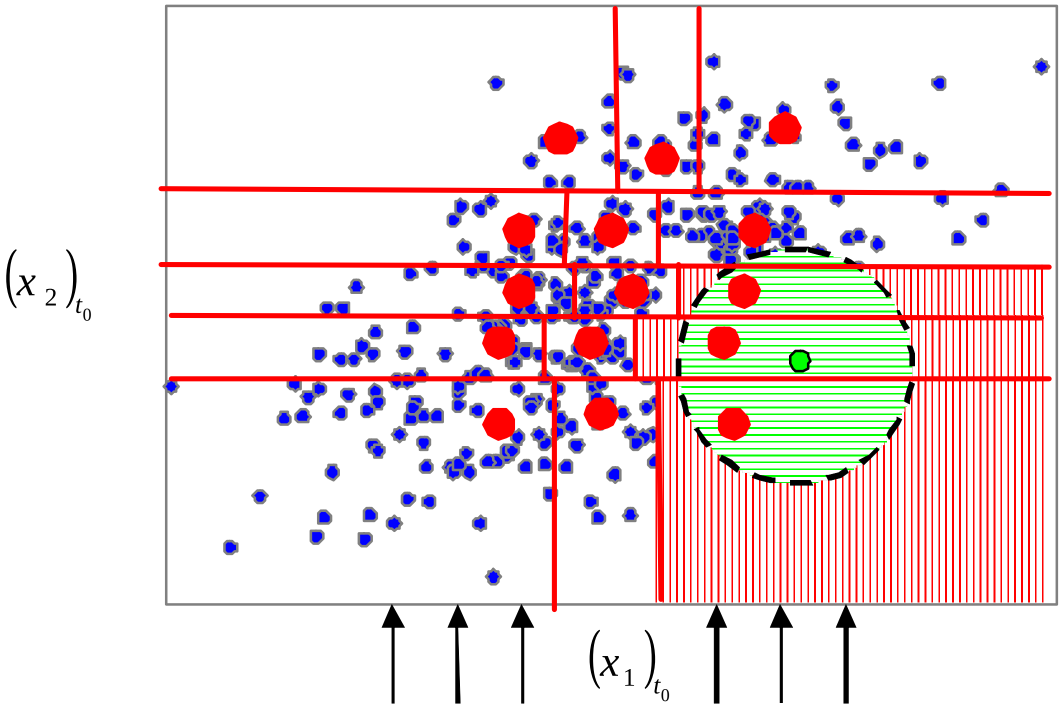
$$p(\mathbf{y}_{t+\tau}|\mathbf{x}_t) = \iint \ldots \int p(\mathbf{y}_{t+\tau}|\mathbf{x}_{t+\tau-1})p(\mathbf{x}_{t+\tau-1}|\mathbf{x}_{t+\tau-2})\ldots p(\mathbf{x}_{t+1}|\mathbf{x}_t)d\mathbf{x}_{t+\tau-1}\ldots d\mathbf{x}_{t+1}$$

- Multi-period predictive distributions estimated empirically:
  - Random vector $\mathbf{x}_{t+1}$ simulated from the one-period predictive distribution using re-sampling techniques (bootstrap) or by sampling from an approximating analytical distribution (using copulas or kernel smoothing)
  - Using the simulated values from $\mathbf{x}_{t+1}$, the corresponding values of $\mathbf{y}_{t+2}$ simulated for the next period. 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 $\mathbf{x}_{t+1}$.

Adaptive segmentation approach: Advantages over existing methods

- Works with various shapes of variable distribution & with different correlation structures (no limiting assumptions imposed)
- Full information, contained in the variable probability distribution, preserved
- Unlike Markov Chain & other probabilistic models using customer segments, our model adaptively seeks for homogenous customer segments without loss of information
- The model can work with partial information &missing variables producing a meaningful forecast for new customers
- The effect of errors & outliers in the raw company data is minimized in our scaling

Optimisation of computational efficiency



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

Segment size is the same across all coarse segments to ensure equal representation. Coarse segment size, $\sqrt{N}$, chosen with an objective to minimize the amount of computation.

Model estimation and validation

Objectives of validation:

- choose a set of predictive variables
- evaluate the predictive performance of the model

Model validation uses retail customer data of 467,789 customers of National Australia Bank Europe over 2005-2008

Purpose of study & potential problem:

predict customer revenue as an outcome of customer behaviour BUT: revenue values affected by external factors even if customer behaviour does not change over time, e.g.:

- sales margins change due to change in the BE base rate
- mismatch between the BE rate, costs of wholesale borrowing and interest rates on retail financial products

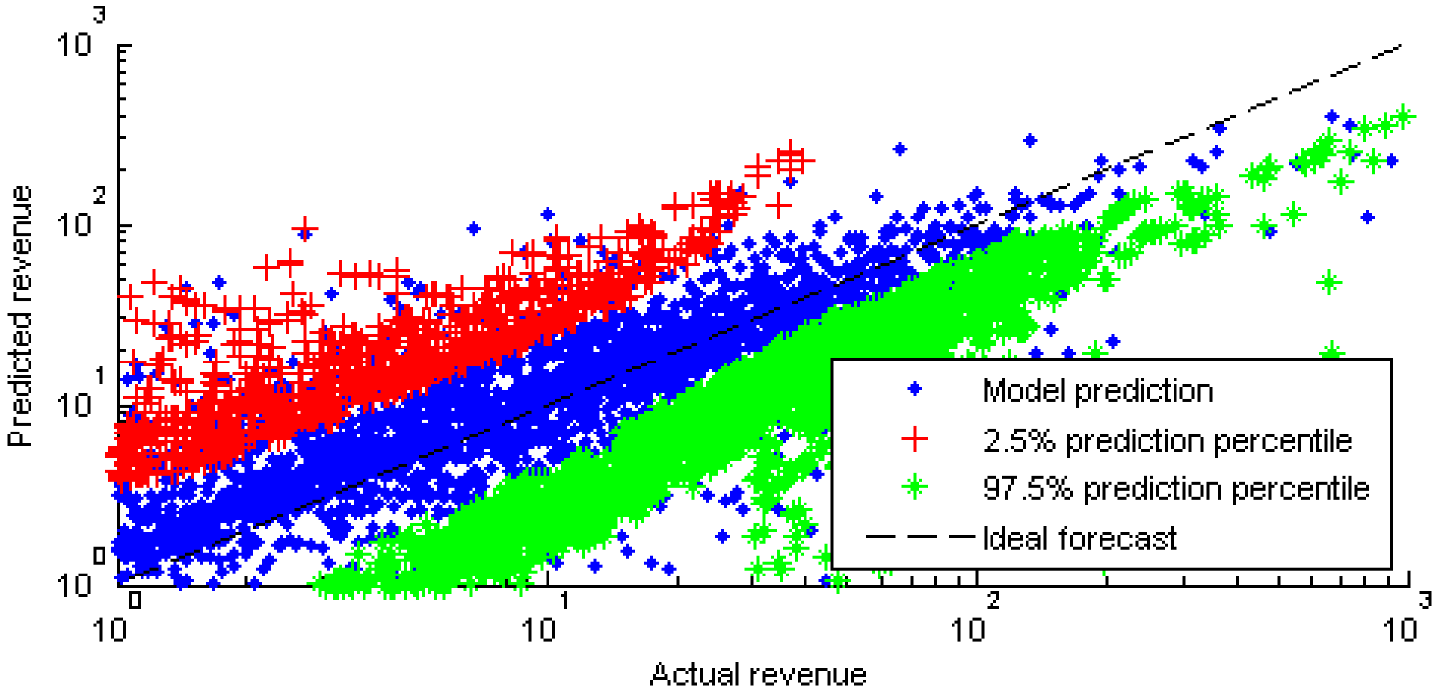
Solution:

Focus on maximizing the rank correlation between the predicted and actual values

- Kendall's tau is insensitive to any variable transformation which does not change the ordering in the population.
- Model preserves the relative ordering in the dependent variable (the revenue from a customer) & 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

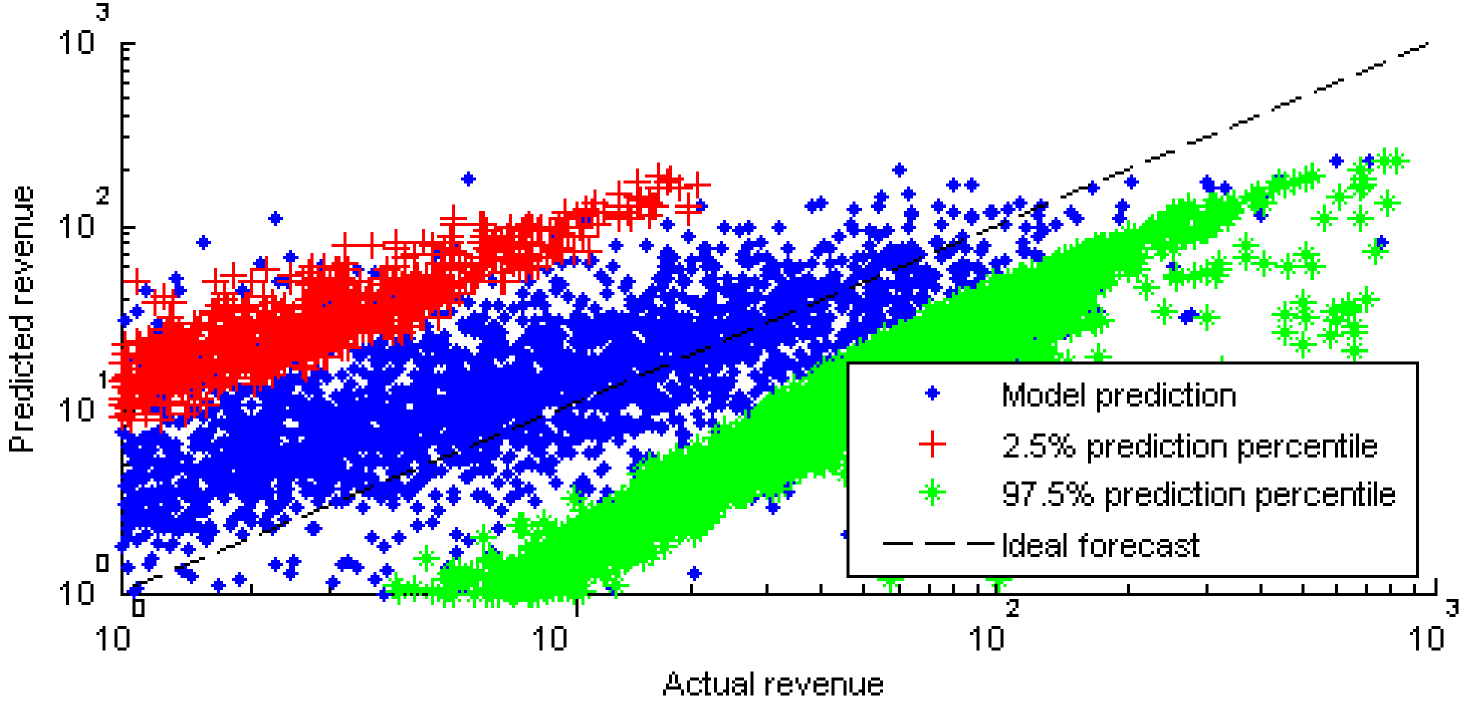
Validation of one-period-ahead forecast

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



Validation of two-period-ahead forecast

Actual versus predicted customer revenue; logarithmic scale
Two-period-ahead prediction



Validation of one and two-year-ahead forecast of customer revenue

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

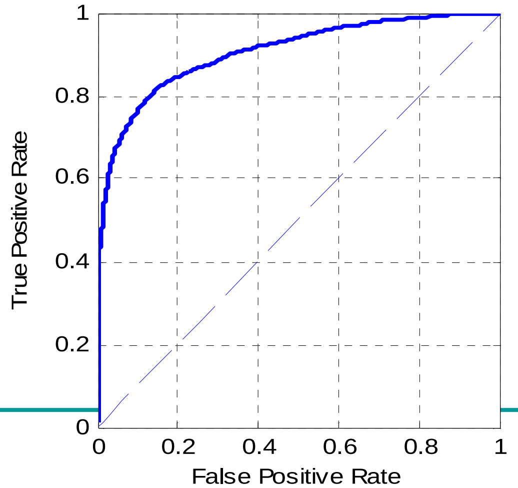
Other applications: predicting large jumps in revenue from a customer

One-period ahead forecast of binomial variables: (1) 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’)

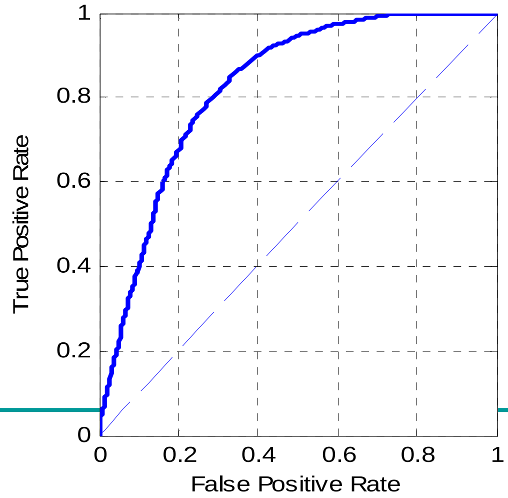
Forecasted variable	Area Under Curve	Accuracy
Stable	0.9068	0.8488
Jump Up	0.8256	0.7566
Jump Down	0.9100	0.8690

ROC curves for the one-period-ahead prediction of customer behaviour

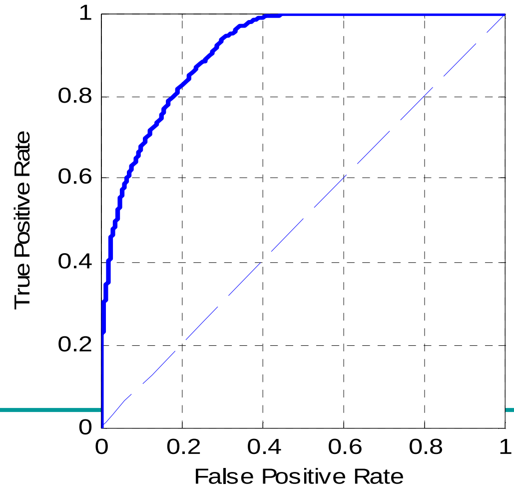
‘Stable’: change in revenue within the threshold



‘Jump up’: increase in revenue above the minimum threshold



‘Jump down’: decrease in revenue below the minimum threshold



Conclusion

Propose an adaptive segmentation approach to the modelling of lifetime value for individual customers in a multiservice financial organisation

- Main Advantages:
  - Model adaptively seeks to locate homogenous customer segments without loss of information about variable distributions compromising the accuracy of prediction
  - Approach does not require assumptions about the shape of the variables distributions or the correlation structure between them;
  - The model can work with partial information and missing values
- Model is validated and implemented for a UK retail bank
- Gives robust predictions using a small number of predictive variables
  - the revenues from individual customers
  - significant changes in revenues
- Validation using 2008 data, coinciding with the crisis year in banking, confirmed the robustness of our relative ranking approach
- Other potential applications: prediction of other customer-related characteristics & behaviour
- Provides a powerful tool in the development of tailored customer management strategies (e.g., customer acquisition and retention)