Cross-Selling Models for Private Banking
(Extended Abstract)

Özden Gür Ali
Koç University, Istanbul, Turkey

Haldun Aytuğ
University of Florida, Gainesville, FL, USA

Hamdi Özçelik
Yapı Kredi Bank, Istanbul, Turkey

Introduction

Cross-selling, offering the right product to the right customer at the right time, is one of the most effective techniques to convert consumers to repeat buyers. The belief is that if a customer is interested in one product there is a high likelihood that she will be interested in others that are somehow related to the current purchase. The main task, of course, is to identify which products are most appropriate to offer next. Enticing customers to adopt new products not only improves the bottom line but also converts new customers to long-term customers. Unlike consumer products, financial products depend very much on the financial savvy and risk tolerance of the customer. In addition, perceptions on financial products vary with economic conditions.

In this work we focus on a decision problem faced by a bank. The management, like all businesses, is interested in improving revenues on their private banking operations while making sure that the customer is served well. One approach they think might help them is cross-selling. In the banking sector cross-selling is especially important since there is evidence that a customer who has purchased several products from a bank is less likely to churn (Poel and Lariviere, 2004). Given the cost of acquiring new customers and the evidence that increasing sales on existing customers is highly profitable, it is in the bank's best interest to offer the right product to the right customer at the right time.

We investigate various cross-selling models, including Markov and logistic regression models. These predictive models vary in terms of their data needs; some can easily be constructed based on the last few transactions while others require sophisticated demographic, financial, economic and even survey data. We try to answer the question "what type of data is sufficient to create a reliable predictive model that can be used for cross-selling financial products?"

Literature Survey

One of the main findings of cross-selling research is that the financial maturity of customers dictates their purchasing behavior. Financial products vary in their level of complexity. For example, checking and savings accounts do not require any financial savvy but stocks, bonds, currency investments as well as insurance products are relatively complex and do require a thorough understanding of risk and return. Li et al (2005) and Kamakura et al. (1991) model this complexity on a continuum and jointly estimate the customers' maturity and products' complexity on the same continuum. The idea is that customers should be offered products that are close to their maturity on this continuum. Kamakura et al. (1991) introduce this single continuum idea and provide a model based on latent trait analysis. Li et al. (2005) find that gender, age, education level do matter in determining how a bank should market their products.

Knott et al (2002) provide a logistic regression model without explicitly measuring the customers' financial maturity. The model predicts the next likely purchase based on customer demographics, recency, frequency and money (RFM) variables as well as current product ownership.

Prinzie and Poel (2006) use a completely different approach relying on sequence analysis. The main idea in this approach is that the sequence in which products are acquired provides a pattern representing a customer's maturity. Consequently, customers who have followed a particular pattern are likely to buy similar products. They use Markov Chain models with lags to estimate the likelihood of the next purchase.

Models

In addition to the predictive power of the models, practitioners are concerned with the complexity of the models and the cost of preparing and maintaining the required input data for the model. Our main aim in this work is to explore the complexity – cost – predictive accuracy space and identify a simple model with few inputs that has acceptable predictive power. We use Markov models as base line models since they are the simplest to build. As in Knott et al. (2002) we build more complex logistic regression models using independent variables from various different categories including demographics, current and past product ownership, affinity to certain product groups, product returns, economic indicators at the time of transactions, etc. Some of these variables readily exist in the bank's database some have to be derived.

The data represents monthly customer transactions spanning a period of approximately two years between 2008 and 2010. For that period we also have snapshots of customer portfolios, demographic information, product profiles as well as external information such as inflation, exchange rates, etc. There are roughly ten thousand customers and we display the results for 20 products. We split the data roughly 90% and 10% for model construction and testing respectively. All reported performance measures are on test data.

Below we summarize model details and our findings.

Markov Models

As in Prinzie and Poel (2006) we built models that use the last purchase as well as last few purchases. That is, given *m* products and purchase histories we estimate

$$P\left\{X_{t+1}=j \,\middle|\, X_t=i_0,\ldots,X_{t-k}=i_k\right\}$$

where j represents the index of the next product and $i_0, \ldots, i_k$ represent the last purchase, the one before last, etc. For models with more than one lag we use the procedures suggested by Raftery and Tavare (1994) and Berchtold and Raftery (2002) and implemented in Berchtold (1999).

One model, called the mixture transition density (MTD), estimates the desired k step probability as a convex combination of one step transitions. That is,

$P\left\{X_{t+1}=j \middle| X_t=i_0,\ldots,X_{t-k}=i_k\right\}=\sum_{g=0}^{k}\lambda_g P\left\{X_{t+1}=j \middle| X_{t-g}=i_g\right\}=\sum_{g=0}^{k}\lambda_g P_{i_g j}$. A generalization called MTDg allows associating each lag with a different transition matrix, yielding

$P\left\{X_{t+1}=j \middle| X_t=i_0,\ldots,X_{t-k}=i_k\right\}=\sum_{g=0}^{k}\lambda_g P\left\{X_{t+1}=j \middle| X_{t-g}=i_g\right\}=\sum_{g=0}^{k}\lambda_g P_{i_g j}^{(g)}$ where $P_{i_g j}^{(g)}$ and λ_g are the transition probabilities and weights associated with lag g respectively.

Product	MC-1	MTD-2	MTDg-2	MTD-3	MTDg-3
A	0.68	0.69	0.63	0.69	0.67
B	0.73	0.70	0.73	0.72	0.74
C	0.74	0.69	0.73	0.74	0.73
D	0.77	0.75	0.76	0.75	0.76
E	0.78	0.80	0.80	0.81	0.79
F	0.76	0.79	0.81	0.82	0.80
G	0.65	0.65	0.66	0.66	0.67
H	0.66	0.66	0.67	0.66	0.69
I	0.61	0.63	0.64	0.62	0.64
J	0.60	0.65	0.65	0.71	0.77
K	0.66	0.69	0.70	0.69	0.70
L	0.63	0.60	0.63	0.59	0.60
M	0.79	0.80	0.81	0.77	0.78
N	0.68	0.61	0.60	0.63	0.65
O	0.66	0.63	0.67	0.66	0.68
P	0.67	0.67	0.67	0.68	0.68
Q	0.66	0.74	0.80	0.72	0.79
R	0.70	0.66	0.67	0.64	0.66
S	0.65	0.63	0.66	0.65	0.66
T	0.73	0.60	0.66	0.59	0.61

Table 1- Area under the curve for Markov models

We report findings on models with up to lag three. We measure model accuracy based on whether the model predicts the next product to purchase correctly and report the area under the receiver operating characteristic (ROC) curve. We refer to this measure as area under the curve (AUC). AUC values for twenty products, based on test data, are reported in Table 1. We see that in general MTDg models are slightly better and both the two-lag and three-lag models perform the same on average. We observe, however, differences in performance based on the product of interest; an issue that needs to be investigated further.

Logistic Regression Models

The logistic regression models predict whether the customer's portfolio will include the product of interest in the next month, given that it is not present in the current month's portfolio. The Formally

$$y_{j(t+1)p} = \begin{cases} 1 & \text{if customer } j \text{ purchases product } p \text{ at time } t+1 \\ 0 & \text{if customer } j \text{ does not purchase product } p \text{ at time } t+1 \end{cases}$$

The logistic regression model expresses $y_{j(t+1)p}$ as a function of x_{jtp} which is a vector of predictor variables. The logistic function is given by

$$ln\left(\frac{p_{j(t+1)p}}{1-p_{j(t+1)p}}\right) = \beta' x_{jtp} + \varepsilon_{jtp}$$

where β' is the transpose of the coefficient vector. Here ε_{jtp} is the error term assumed to be i.i.d. with logistic distribution. β is a vector consisting of coefficients for the predictor variables.

The bank has access to seven different sets of variables. Some, such as product use, are easy to extract whereas others do require significant manipulation of customer data, for example affinity-scores. All are available internally with the exception of Economic-indicators covering the span of time for which the models are built. In this analysis we used five of these sets of variables which we list below.

- Products in customer's current portfolio (Current-use)
- Whether the customer has ever used the product before (Ever-used)
- Demographics (age, tenure, gender, nationality, education, occupation, marital status, etc.)
- Monthly product return percentages for the current month (Returns)
- Economic indicators (e.g., consumer_Confidence, TUFE_12month_inflation, IMKB100, etc.)

We report the AUC values of twenty products based on several models in Table 2. It is clear that the model that uses all five information sets performs best. However, we also observe that models built on variables indicating current and/or past ownership of products (columns ever-used, current_use) are serious competitors. Work is currently under way to determine the smallest information set that has performance comparable to that of the last column in Table 2.

Product	Current-use	Ever-used	Demographics	Returns	Economic Indicators	Current_use + Ever_used	Cur_use + Ever + Dems + Returns + Economic
A	0.68	0.75	0.61	0.79	0.75	0.76	0.90
B	0.70	0.71	0.56	0.78	0.76	0.75	0.84
C	0.65	0.68	0.55	0.80	0.79	0.69	0.84
D	0.71	0.71	0.50	0.77	0.73	0.73	0.84
E	0.69	0.73	0.48	0.68	0.65	0.73	0.77
F	0.74	0.74	0.58	0.74	0.71	0.77	0.85
G	0.78	0.80	0.54	0.61	0.61	0.82	0.83
H	0.85	0.90	0.54	0.52	0.39	0.89	0.90
I	0.87	0.89	0.68	0.64	0.47	0.85	0.86
J	0.67	0.69	0.54	0.63	0.61	0.71	0.74
K	0.80	0.84	0.59	0.55	0.60	0.84	0.84
L	0.64	0.72	0.52	0.62	0.60	0.72	0.79
M	0.79	0.82	0.45	0.62	0.52	0.81	0.83
N	0.73	0.74	0.58	0.59	0.54	0.77	0.78
O	0.77	0.81	0.64	0.63	0.56	0.83	0.86
P	0.73	0.79	0.64	0.67	0.61	0.81	0.85
Q	0.84	0.84	0.50	0.82	0.78	0.90	0.94
R	0.67	0.75	0.58	0.52	0.51	0.77	0.79
S	0.65	0.71	0.59	0.61	0.62	0.72	0.76
T	0.69	0.74	0.62	0.65	0.62	0.74	0.79

Table 2 - Area under the curve for various logistic regression models

Conclusion and Future Research

In this work we explore variables expressing customer's current and past use of different product groups, product characteristics, demographics and economic indicators. Markov based sequence analysis shows that the most recent purchase is an effective predictor and looking back a few purchases improves predictive accuracy for some products. Models that use additional information, such as logistic regression models, improve predictive accuracy. Models built on demographic information alone or product performance alone do not perform well. Interestingly, models that use only current or past ownership (i.e., whether product *x* was ever purchased) do perform well. Combining purchase history, product performance, demographics and economic indicators have the highest predictive accuracy for almost all products (with few exceptions that are relatively easy to explain).

On the research side, work remains on identifying the most parsimonious model and incorporating the results of a customer survey. On the implementation side, which models will ultimately be used will depend on factors such as model parsimony, acceptance by users (account managers), costs of data collection in addition to predictive accuracy.

References

1. Berchtold, A, “GMTD version 1.0, User’s Guide” University of Washington, Department of Statistics, Seattle, WA, (1999) pp.1-24.
2. Berchtold, A. and Raftery A. E. "The Mixture Transition Distribution Model for High-Order Markov Chains and Non-Gaussian Time Series" *Statistical Science* (17) 2002 pp. 328-356.
3. Kamakura, W. A., Ramaswami S. N. and Srivastava R. K. "Applying latent trait analysis in the evaluation bf prospects for cross-selling of financial services" *International Journal of Research in Marketing* (8) 1991 pp. 329-349.
4. Knott, A., Hayes A. and Neslin S. A. "Next-Product-toBuy Models for Cross-Selling Applications" *Journal of Interactive Marketing* (16:3) 2002 pp. 59-75.
5. Li, S., Sun B. and Wilcox R. T. "Cross-Selling Sequentially Ordered Products: An Application to Consumer Banking Services" *Journal of Marketing Research* (42) 2005 pp. 233-239.
6. Poel, D.V. and B. Lariviere. “Customer attrition analysis for financial services using proportional hazard models” *European Journal of Operational Research*, (157), 2004, pp. 196-217.
7. Prinzie, A. and Poel D. V. "Investigating purchasing-sequence patterns for financial services using Markov MTD and MTDg models" *European Journal of Operational Research* (170) 2006 pp. 710-734.
8. Raftery, A. E. and Tavare S. "Estimation and Modeling Repeated Patterns in High Order Markov Chains with the Mixture Transition Distribution Model" *Applied Statistics* (43) 1994 pp. 179-199.