Paper 163

Behind Marketing Data

Flavio Addolorato[1] and Silvia Pacei[2]

[1] Economic Research Department, Banca Commerciale Italiana, Via Borgonuovo, 2, 20141 Milano, Italia.
[2] Department of Statistics „P. Fortunati“, Università di Bologna, Via Belle Arti, 41, 40126 Bologna, Italia.

Summary

The aim of the work is to pinpoint a number of behavioural characteristics of the Banca Commerciale Italiana (BCI) retail customers by using several analytical tools. A panel sample is maintained which is observed each semester. A factor analysis and a cluster analysis are carried out in order to subdivide customers into segments according to the main factors affecting their behaviour. Repeating the same clustering procedure every semester, it is possible to carry out a migration analysis between segments. For this purpose, SAS/Insight is used to explore data through graphs and analyses linked across multiple windows. Moreover, the transitions from one cluster to another are studied following a model-based approach, but the present availability of only three waves makes our assumptions too simple to describe the customers' migration process.

Keywords: Multivariate data analysis, SAS/Insight, Mover-Stayer Model

1 Introduction.

The object of this project is to test the use of several analytical tools in order to pinpoint a number of behavioural characteristics of the Banca Commerciale Italiana (BCI) retail customers.
From January 1999 onward the BCI is going to maintain a panel sample which is observed monthly. This survey will enable us to collect lots of information on bank customers and on their behaviour, but at the moment only the traditionally used semester frequency data is available
A behavioural segmentation analysis is carried out in order to subdivide customers into segments according to the main factors affecting their behaviour. The identification of those segments enables to better understand and meet the requirements of bank customers. Then, repeating the assigning process for every customer to a segment for every semester, it is possible to carry out a migration analysis between segments, as well as to observe the effects of the marketing activity of the bank. For this purpose, SAS/Insight allows us to interactively analyse the migration flows of customers among behavioural segments. This software is useful to interactively explore data through graphs and analyses linked across multiple dimensions.
Moreover, we are now trying to study the transitions from one cluster to another following a model-based approach, which is rarely taken into consideration in studies on bank customers. The main advantages of that kind of approach would be the possibility of describing the migration phenomenon through a model with some particular characteristics and of predicting future customer migrations among groups. The basic assumption of the choosen model is that movements into and out of the segments can be described by a modified Markov process, that is a Mover-Stayer model.

2 The behavioural segmentation of retail customers.

A stratified sample of about 100,000 bank customers is observed each semester from June 1997 onward, in order to get information on their behaviour.
The data was sourced from the corresponding operative product files and then reclassified on the basis of a „*phisycal person client* “ logic. In the case of joint account relationships, we decided to duplicate the products utilized for each of the customers involved. A considerable amount of information is available (the length of customers’ relationships with the bank, quantitative characteristics of the bank products they own, the frequency of the operations done on their accounts, etc) and a factor analysis is carried out, in order to explain the correlation between these variables in terms of a smaller number of underlying factors (Jobson, 1992).
Starting from information regarding only current accounts and securities deposits, the factor analysis enables us to identify six standardised and uncorrelated factors,

highly representative of the overall variability. By observing the estimates of the coefficients connecting factors and original variables, it is possible to attribute an economic meaning to each factor. The first factor seems to represent the typical transactions of retail customers (who mainly use cheques), while the second factor seems to represent professional customers (who mainly use allowances). The third and the fourth factors are connected to the loyalty to the bank and the last two seem to represent the frequency and the amount of transactions.

On the basis of these six factors, a first segmentation was carried out using a cluster analysis of hierarchical type on a sub-sample in order to identify the best possible number of segments; later a non-hierarchical cluster analysis was performed to finalise groups. The process resulted in the identification of six segments, one of them in particular comprised a large number of clients (47% of the sample) lacking specific distinctive features. It was decided to experiment a neuronal classification technique called Kohonen Map (Kohonen, 1995) only on the clients belonging to this segment. This enabled to identify eleven segments altogether, through the use of a Multi Layer Perceptron-type neuronal network, to work out an algorithm to assign customers to clusters on the basis of the initial behavioural variables. Through this algorithm, in the three considered semesters, it was possible to classify retail customers into segments maintaining the same logical meaning.

Correspondence analysis, another multivariate analysis technique which allowed us to represent the relationships between a set of qualitative variables on a two-dimensional plane, is used to help describe segments. For this purpose, other information on sample units is introduced in the analysis, such as demographic characteristics and the qualitative characteristics of their bank products.

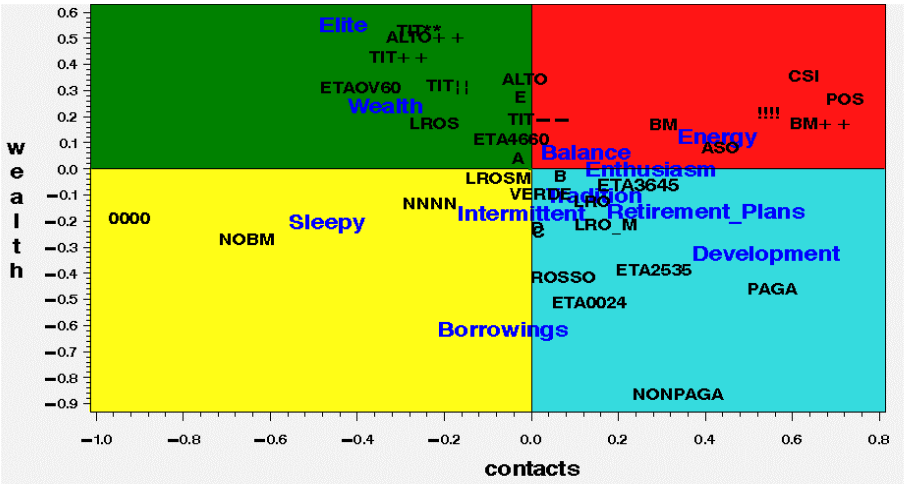

Fig. 1

In the diagram (Fig. 1) the cluster positions and the corresponding variables allow us to interpret the two dimensions as wealth and frequency of contacts (that is level of customers' activity). Similar segments have been found by other studies on bank customers behaviour (Addolorato et al., 1998). In the diagram we can find the labels by which we try to interpret the distinctive features of each segments: ***Tradition,*** these customers probably appreciate a personalized relationship with a branch, while it does not seem they would be a well suited target for high technology products; ***Wealth,*** the customers in this group would appear to be natural candidates for institutionally managed savings products. In this case, the bank could also propose some technologically sophysticated services; ***Enthusiasm***, this group of young customers is open to technological innovation; ***Development,*** this group could be interested in instruments which to build up capital over time; ***Retirement Plans,*** these customers could be interested in the basic products in the field of institutionally managed saving; ***Intermittent,*** improvement in customer loyalty is needed in this segment, and could be achieved, for example, with efforts to facilitate the use of bank cards; ***Borrowings,*** starting from the outstanding loans, the bank could aim at developing base products with this group*;* ***Sleepy or Poised to Flee*** these are customers to be cultivated; they transact only marginal business with the bank, they could be the target of initiatives to revive relationships; ***Elite,*** these are customers with highly personalised traditional products (essentially, private banking); ***Energy,*** this is fertile ground for innovative products across the entire range of bank services; ***Balance,*** these customers are quite loyal, and open to the possibilities of new products and new types of distribution channels.

3 The interactive graphical analysis of the migration between clusters.

To make our analysis easier to understand, we chose to consider only the migration semester-to-semester flows between the quadrants of the correspondence analysis map (A,B,C,D from top left to clockwise) and to explore them using SAS/Insight (SAS Institute Inc., 1995), with the objective to measure the marketing activity of the bank (Fig. 2).

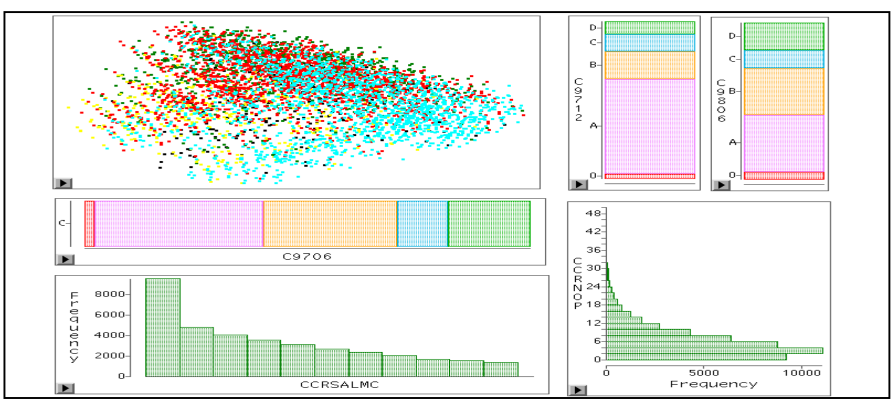

Fig. 2

4 The migration analysis with a Mixed Markov model.

The object of this step of the analysis is to verify if a model as simple as a Mixed Markov process model could be adequate to describe movements between customers' clusters (in this case the possible state are represented by the four quadrants resulting from the correspondence analysis map: ABCD from top left clockwise). We suppose the length of time that customers spend in each state has important influence on the probability of moving during the subsequent period. Besides, it is possible to observe that a certain proportion of each state does not move in the two transition periods taken into consideration. Hence a modified Markov process is assumed, that is a Mover-Stayer model in which some members of the population never move, whereas the others change state following a simple first-order Markov process.

This model-based analysis is carried out leaving out customers that move in or out of the bank. Indeed, gains and losses are an important feature of the studied process, but we believe that they are governed by a stochastic mechanism which is different from the mechanism governing the internal migration process. Our interest thus focussed on the changing internal structure of the system in this first step of the analysis.

In the two considered time intervals customers move in and out of the bank especially from quadrant D: about 40% of the whole loss and 33% of the whole gain belong to it. Hence, D can be considered the weakest and most unstable state, which the bank should turn its marketing efforts to.

Moving back to the theory of the model, the usual first-order Markov chain assumes that movements into and out of each state depend only on one's present position and are independent of all prior history. **P** is the matrix of transition probabilities, whose entry p_{ij} ($i=1,\dots,k;\ j=1,\dots,k$ and k is the number of the states) represents the conditional probability of going to *j* in the next period given that one is currently in *i*. The law of motion for the assumed Mover-Stayer process is given by:

$$p_{ij}=\begin{cases} s_i+(1-s_i)m_{ii}, & i=j=1,\dots,k, \\ (1-s_i)m_{ij}, & i\neq j. \end{cases}$$

In this model the proportions of stayers in each segment are given by s_i ($i=1,\dots,k$); m_{ij} ($i=1,\dots,k;\ j=1,\dots,k$) are the elements of the matrix of transition probabilities for the movers $\mathbf{P}_M$, which is assumed to represent a simple first-order Markov process (Bartholomew, 1982). Assumptions of that model have already been considered plausible above all in economic studies on occupational mobility (Regoli, 1994; Van de Pol and Langeheine, 1992).

Information on customers' characteristics are typically observed every semester and only three semester segmentations are available for this study. Since the characteristics of the bank products that customers own as well as customers' behaviour usually change at the end of the year, there will possibly be a seasonal effect on the segmentation obtained for the three considered waves. Hence we expect that the transition probability in $\mathbf{P}_M$ are not time-homogeneous (non stationary process) and two different transition probabilities matrixes for the movers have to be estimated.

Maximum likelihood estimates of the transition probabilities and of the proportions s_i are obtained with the EM algorithm both assuming a stationary and a non-stationary process. The maximum likelihood estimates of the standard errors of these parameters are calculated as well. Comparing the effectiveness of the two models by looking at the likelihood ratios (G^2 = 863 with 32 degrees of freedom for the non stationary model and G^2 = 7,209 with 44 degrees of freedom for the stationary one), it turns out that the fit looks really better for the non stationary model, as it was expected, even thought it may not be considered very satisfactory in this model as well. Standard errors of the chain proportions estimates are quite low as well as standard errors of the transition probabilities estimates in the non stationary case (in average 3% of estimated values).

Evidently, the simple assumptions underlying the selected model are not completed suitable to describe the studied process. Anyway, it is not to be excluded that a model based on more complex assumptions may be adequate. When further waves are available it will be possible to estimate the model by using annual data, that is information not affected by seasonal factors, and/or to assume a more complex stochastic mechanism governing the movers migration process like, for instance, a second-order Markov chain.

References

Addolorato F., Banfi , Bodio L. (1998), *Profili di segmentazione comportamentale della clientela: un case study*, in: Banfi A., Di Battista M.L., Tendenze e prospettive del risparmio gestito, il Mulino, Milano

Bartholomew D. J. (1982), *Stochastic Models for Social Processes*, John Wiley & Sons: New York.

Jobson J. D. (1992), *Applied Multivariate Data Analysis*, Springer-Verlag: New York.

Kohonen T., (1995), Self-Organizing Map, Berlin, Springer.

Pol F. J. R. van de, Langeheine R. (1992), *Analysis Measurement Error in Quasi-Experimental Data: an Application of Latent Class Models to Labour Market Data*, in "Working Papers of the European Scientific Network on Household Panel Studies", Paper 57, University of Essex, Colchester.

REGOLI A. (1994), "L'analisi delle transazioni nel mercato del lavoro in presenza di errori di classificazione", *Economia & Lavoro*, anno XXVIII, n. 3-4.

SAS Institute Inc. (1995), *SAS/INSIGHT User's Guide*, Version 6, Third Edition, Cary, NC: SAS Institute Inc.