Longitudinal models for dynamic segmentation in financial markets

Francesca Bassi, Department of Statistical Sciences, University of Padova, Italy

Introduction

In this paper, two alternative approaches to clustering categorical data are compared: the latent class Markov (LCM) approach and the latent class growth (LCG) approach. Markov type models describe change between consecutive time points while growth models study the development of behaviors over time. Both approaches have been shown to be very interesting for dynamic market segmentation (Paas *et al.*, 2007, Bassi, 2012). In this paper the focus is on understanding which approach performs better in terms of fit and of usefulness of the obtained results. The LCM model represents dynamics over time across latent states with conditional probabilities following a Markov chain. Covariates may affect both the initial state of the latent chain and transition probabilities. In the context of market analysis, latent states aim at representing segments of consumers. The LCG model identifies distinctive groups with different trajectories: it may be seen as an extension of the growth curve model (Nagin, 2005) in which a categorical variable is used to capture heterogeneity in developmental trajectories; moreover, the relationship between latent classes and covariates may be estimated as well as the effect of covariates on the latent trajectory. Groups identified by the mixture may be interpreted as market segments.

The topic of market segmentation is still one of the most pervasive in marketing. The overall goal of segmentation is to divide a population into mutually exhaustive and exclusive subgroups which differ with respect to some criteria and to identify those segments which are the best from a marketing perspective so that they can be targeted. Moreover, in order to design appropriate marketing strategies, other information are fundamental (Prinzie and van der Poel, 2011).

There is a rich literature on segmentation in the financial market showing that this a relevant instrument to design marketing strategies in the field. Most studies are based on classical statistical instruments such as cluster analysis and on data collected on convenience samples of individuals. Patsiotis *et al.* (2012), for example, cluster a convenience sample of customers in order to profile adopters and non-adopters of internet banking; Machauer and Morgner (2001) propose to identify segments in the financial market through variables describing benefits and attitudes of a convenience sample of individuals. A more innovative work is that by Zuccaro and Savard (2010), who propose an hybrid segmentation of a sample of clients of a large bank linking their profiles to the segments identified by the Canadian Mosaic segmentation portal. Other recent interesting contributions are those by Rink at al. (2015) and Lees et al. (2016).

This paper introduces in the financial market the topic of dynamic segmentation. It is, in fact, important to understand if and how customers move across segments over time. Then, it is necessary to estimate the effect of potential covariates, the characteristics of the household, on the probability to belong to each segment and on that to move from one segment to another over time.

The paper aims at comparing these two approaches, the LCM and the LCG, underlining the specificities of both of them and the type of applied problems that they can handle, possibly suggesting in which practical situation one or the other model should be preferred for performing market dynamic segmentation.

As an example, the models are estimated on data collected by the Bank of Italy with the Survey on Household Income and Wealth on a representative sample of Italian families. Ownership of 13 financial products from 2002 to 2010 and the characteristics of the households referring to the same period of time are considered.

The paper is organized as follows. Section 1 introduces latent class Markov models and Section 2 latent class growth models. Section 3 compares the two approaches, describing how each one may be suitable for dynamic market segmentation and, eventually, in which situation one should be preferred to the other. Section 4 briefly describes the survey and the data used as an example. Section 5 compares results obtained with the two approaches and Section 6 concludes.

1.Latent class Markov models

Let us consider the simplest formulation of latent class Markov models (Wiggins, 1973), which assumes that true unobservable transitions follow a first-order Markov chain. As in all standard latent class model specifications, local independence among the indicators is assumed, i.e., indicators are independent conditionally on latent variables[1].

Let X_{it} denote segment belonging at time *t* for a generic sample unit *i*, *i=1,..,n*; Y_{ijt} is an observed categorical variable related to item *j, j=1,…, J* for unit *i* at time *t*; $P(X_{i1}=k_1)$ is the probability of the initial state of the latent Markov chain, and $P(X_{it}=k_t / X_{it-1}=k_{t-1})$ is the transition probability between state k_{t-1} and state k_t from time *t-1* to *t*, with *t*=2,…,*T*, where *T* represents the total number of consecutive, equally spaced time-points over which a unit is observed. Besides, let $P(Y_{ijt}=h_t / X_{it}=k_t)$ be the probability that unit *i* gives answer h_t at time *t*, given that unit *i* at time *t* belongs to segment k_t, this is also called the model measurement component.

For a generic sample unit *i*, a LCM model is defined as:

$$P(\mathbf{Y_{i1}},\ldots,\mathbf{Y_{iT}})=\sum_{k_1}^{K}\ldots\sum_{k_T}^{K}P(X_{i1}=k_1)\prod_{t=2}^{T}P(X_{it}=k_t\mid X_{it-1}=k_{t-1})\prod_{j=1}^{J}\prod_{t=1}^{T}P(Y_{ijt}=h_t\mid X_{it}=k_t) \tag{1}$$

where

$\mathbf{Y_{i}}_t$, is the vector containing the values of the observed variables, or indicators, at time *t* for unit *i*, k_t varies over *K* latent states and h_t over a set of *H* categories.

In a LCM model with concomitant variables, latent state membership and latent transitions are expressed as functions of covariates with known distribution (Dayton and McReady, 1988): $P(X_{i1}=k_1/\mathbf{Z}_{i1}=\mathbf{z}_1)$, where $\mathbf{z}_1$ is a vector containing the values of covariates for unit *i* at time *1*, estimates covariates effects on the initial state and $P(X_{it}=k_t/X_{it-1},\mathbf{Z}_{it}=\mathbf{z}_t)$, where $\mathbf{z}_t$ is a vector containing the values of covariates for household *i* at time *t*, estimates covariates effects on latent transitions.

[1] In the LCM model with one indicator per latent variable, the assumption of local independence coincides with the Independent Classification Error (ICE) condition.

On the bases of the above defined components, the complete model for unit i is given by:

$$P(\mathbf{Y}_i = \mathbf{y} \mid \mathbf{Z}_i = \mathbf{z}) = \sum_{k_1}^{K} \ldots \sum_{k_T}^{K} P(X_{i1} = k_1 \mid \mathbf{Z}_{i1} = \mathbf{z}_1) \prod_{t=2}^{T} P(X_{it} = k_t \mid X_{it-1} = k_{t-1}, \mathbf{Z}_{it} = \mathbf{z}_t) \prod_{j=1}^{J} \prod_{t=1}^{T} P(Y_{ijt} = h_t \mid X_{it} = k_t)$$

where

$\mathbf{Y_i}$ is the vector containing the values of the observed variables for unit i in the T measurement occasions,

$\mathbf{Z_i}$, is the vector containing the values of the covariates for unit i in the T measurement occasions.

Typically, conditional probabilities are parameterized and restricted by means of logistic regression models.

Parameter estimation is performed via maximum likelihood using the E-M algorithm (Dempster *et al*, 1977).

2.Growth models

Growth models study the development of individuals over time capturing the dependence introducing one or more latent variables (Vermunt, 2007). Basically, growth models are regression models for two-level data – time points nested within individuals – in which time enters as a predictor.

Let Y_{it} be an observed variable denoting response for unit i at time t, with $i=1,\ldots,n$, and $t=1,\ldots T$, equally-spaced time points. These repeated observations are regarded as imperfect measures of an underlying latent trajectory. The shape of the growth trajectory (linear, quadratic, etc.) depends on the number of latent variables specified in the model as well as on how the loadings of these latent variables change with respect to time and can be described by the following equation:

$$y_{it} = \alpha_i + \gamma_t \beta_i + \varepsilon_{it}$$

with random effects – intercept and slope - given by:

$$\alpha_i = \mu_\alpha + \zeta_{\alpha_i} \text{ and}$$
$$\beta_i = \mu_\beta + \zeta_{\beta_i}.$$

The model assumes $\zeta_{\alpha_i} \sim N(0, \Psi_\alpha)$, $\zeta_{\beta_i} \sim N(0, \Psi_\beta)$ and $\varepsilon_{it} \sim N(0, \Theta_t)$. ζ_{α_i}, ζ_{β_i} and ε_{it} are mutually independent for every i and t. A usual convention for linear growth is that $\gamma_t = t - 1$. The parameters of interest are the means and variances of the random effects and the residual variances over time. This model is defined an unconditional (with no predictors of growth) growth model. Conditional growth models, instead, not only describe but also explain growth examining predictors of individual change over time. Predictors may be time-constant or time-varying. The random effects are, consequently, specified as

$$\alpha_i = \mu_{0\alpha} + \boldsymbol{\mu}_{1\alpha}\mathbf{x}_i + \zeta_{\alpha_i} \text{ and}$$
$$\beta_i = \mu_{0\beta} + \boldsymbol{\mu}_{1\beta}\mathbf{x}_i + \zeta_{\beta_i},$$

where $\mathbf{x}_i$ is a vector containing the value of time-invariant explanatory variables for individual *i*.

The latent class growth model assumes that the population is heterogeneous, and different subpopulations are characterized by different trajectories (Connel and Frey, 2006). The model estimates the intercept and the slope for each class and individual variation around these growth factors:

$$\alpha_{il} = \mu_{0\alpha l} + \boldsymbol{\mu}_{1\alpha l}\mathbf{x}_i + \zeta_{\alpha_{il}} \text{ and}$$
$$\beta_{il} = \mu_{0\beta l} + \boldsymbol{\mu}_{1\beta l}\mathbf{x}_i + \zeta_{\beta_{il}},$$

where *l* indicates one of the *L* subpopulations with probability π_l.

When indicators are not continuous, after an appropriate transformation, the expected value of the response variable is assumed to be a linear function of a set of predefined functions of time (Vermunt, 2007). In the case of binary response variables, for example, the logit transformation is used.

3. Latent class Markov and latent class growth models compared

Both approaches described in the previous paragraphs can be usefully employed for dynamic segmentation since they allow to identify groups of customers with similar characteristics, i.e., market segments, to follow their behavior over time and, if necessary, to predict it. However, the two approaches rely on different assumptions.

Growth mixture models may be seen as an extension of growth curve models, they aim at describing and testing hypotheses about between-person differences and within- person change. A latent categorical variable, the mixture component, is used to take into account the heterogeneity in the observed developmental trajectories, in this way, the population variability in growth is modeled through a mixture of differently distributed subpopulations. The aim is to get a reliable estimate of the shape of trajectories, of class probabilities, as well as of variation in classes. The conditional latent class growth model estimates the effects of potential covariates on the random parameters of the latent trajectories. Covariates' effects on classes may be also estimated. Units are assigned to a latent class over the observational period, changes across classes is not allowed, a common trajectory is modeled for all individuals belonging to a class.

Latent class Markov models are mainly used for the analysis of categorical longitudinal data and their main feature is that the individual characteristics of interest, and their evolution over time, are represented by a latent process with state occupation probabilities which are time-varying according to a first-order Markov chain. The available covariates may affect the conditional distribution of the response variables given the latent process (the measurement component of the model) and\or the distribution of the latent process (the structural component of the model). Estimated initial probabilities of the chain identify latent groups of units, while estimated transition probabilities describe patterns of changes across latent states over time, therefore units may change the latent state to which they belong over the observational period.

It is interesting to recognize with Vermunt *et al*. (2008) that the latent class Markov model and the latent class growth model are special cases of the mixture latent Markov model:

$$P(\mathbf{Y}_i = \mathbf{y} \mid \mathbf{Z}_i = \mathbf{z}) = \sum_{l=1}^{L}\sum_{k_1=1}^{K}\ldots\sum_{k_T=1}^{K} P(W = l \mid \mathbf{Z}_{i1} = \mathbf{z}_1) P(X_{i1} = k_1 \mid W = l, \mathbf{Z}_{i1} = \mathbf{z}_1) \prod_{t=2}^{T} P(X_{it} = k_t \mid X_{it-1} = k_{t-1}, W = l, \mathbf{Z}_{it} = \mathbf{z}_t)$$
$$\prod_{j=1}^{J}\prod_{t=1}^{T} P(Y_{ijt} = h_t \mid W = l, X_{it} = k_t, \mathbf{Z}_{it} = \mathbf{z}_t) \qquad (2)$$

were W is a time-constant latent variable with L latent classes and all other symbols have the same meaning as in equation (1). This model is composed of three elements, that capture, respectively, unobserved heterogeneity by means of the latent variable W, autocorrelations by means of transition probabilities, and measurement error by means of the relationship between the time-varying latent variables X_t and their indicators Y_t. Moreover, effects of time-constant and time-varying covariates on the three model components may be estimated.

A latent class Markov model is obtained from equation (2) eliminating latent variable W, i.e., assuming that there is not unobserved heterogeneity. The latent class growth model derives from equation (2) eliminating the transition structure and the measurement component; in this case, the time-constant latent variable W captures the dependencies between measurement occasions.

In the application of these models to perform dynamic segmentation some other considerations are important.

(i) The measurement component of the latent class Markov model does not describe misclassification but, as in latent class cluster models (Magidson and Vermunt, 2002), it is used to identify groups of costumers with similar characteristics observed with the indicators. In these applications, latent states, i.e., the categories of the time-varying latent variables, represent market segments. Usually, the measurement component is assumed time-constant in order to ensure that market segments do not change over time.

(ii) In the latent class growth model, categories of the time-constant latent variable, latent classes, identify market segments, i.e., groups of customer that show the same evolution over time with reference to the behavior in the market under study. Customers are not allowed to move across segments in the observational period.

4. The application: the survey and the data

For the application, data on financial product ownership by Italian households collected with the Survey on Household Income and Wealth (SHIW) conducted by the Bank of Italy is used. The survey has been running since 1965 and, with few exceptions, was conducted on an yearly basis till 1987. In the sequel it was run every two years. In 1989 a panel component was introduced so that at each wave the sample consists of two components: a panel sub-sample made up of households who participate in previous waves and a fresh cross-sectional sub-sample. Table

1 contains the survey design and the dimension of the two sample components from 1987 to 2010. The survey collects information on income, savings, consumption expenditure and the real wealth of Italian households, as well as on household composition, demographic characteristics and labour force participation (Giraldo *et al.*, 2001). In the paper, the sample of 1,834 households who participated in all waves from 2002 to 2010 is considered.

Table 1 about here

Information on 5 equally-spaced (two years) time points on ownership of 13 financial products and on family characteristics such as geographical area, gender, age and educational qualification of the head of household is used. Previous studies by the Bank of Italy show that financial activities diffusion among Italian families vary with the selected covariates (Bank of Italy, 2012). Table 2 lists the percentage of households holding, in the five measurement occasions, the 13 financial activities: certificates of deposits (CD), repos (CT), post office certificates (BFP), Treasury bills up to one-year maturity (BOT), fixed-rate long term Treasury bonds (BTP), bonds (OBB), mutual funds (QFC), individually managed portfolios (GP), foreign securities (TE), loans to cooperatives (COOP) and bank or postal deposits (DEP), shares and other equities (SHA), floating rate Treasury certificates indexed to BOTs (CCT). In the period, the incidence of ownership of deposits rose till 2008: between 2002 and 2008 families owning a bank or a postal deposit increased by 2.7%, from 2008 to 2010, they decreased by 3.3%. Our observational period includes the year 2008 when the effects of the economic crisis, in Italy, started to be appreciable. The proportion of households owning government securities fell. From the Italian Ministry of Economics and Finance, we know that the average yield of state bonds at issuance declined from 4.09% to 2.10%. Over the same time span the percentage of households owning shares, investment funds and other risky assets (except for bonds) also declined.

Table 2 about here

Table 3 reports the relative frequency distribution of households for the covariates over the observational period. Obviously, we observe a change in the distribution of heads of household by age and a consequent increase in the percentage of female heads of household. It will be interesting to see if and how the above household characteristics influence financial assets ownership and if changes in household structure determine changes in consumer behavior. The distributions of financial assets by family characteristics from 2002 to 2010 show that financial strategies depend on household structure and socio-economic environment and an interaction between ownership, household situation and time, suggesting an analysis of the market in terms of dynamic segmentation.

Table 3 about here

5. Results

The best fitting LCM model to our data has 5 latent states which represent market segments, a first-order stationary Markov chain and all considered covariates significantly affect the initial state and transitions of the latent chain. In this application, response variables are the 13 binary indicators y_{ijt} assuming value *1* if household *i* holds financial product *j* at time *t*, *i* varies over *1* to *n*, sample dimension, *j* over 1 to 13 and *t* over 1 to 5, since 5 survey waves are considered, from 2000 to 2010. The LCM model with concomitant variables is estimated in one step (Vermunt, 2010) starting with one latent state per each latent variable and increasing the number of latent states till the Bayesian Information Criterion (BIC) increased[2]. On this model, the assumption of time-constant transition probabilities was accepted by means of the conditional test. Response probabilities are set constant in order to make segments stable over time. Table 4 lists estimation results: segment sizes and response probabilities. The largest group (50.10%) is composed of households owning only a bank or a postal deposit and very few forms of other assets, showing that they rely heavily on liquid savings forms for transactional purposes; the group with dimension 12.42% is that comprising the poorest households: they do not own any kind of financial asset, moreover, an important percentage does not even hold a deposit. 19.57% of families owns a deposit and has made some investment in financial assets with a preference for less risky ones such as postal bonds and loans to cooperatives. A segment with dimension 9.23% contains households that mainly possess state bonds (as well as deposits). Finally, a segment made of 8.68% of families owns a deposit, state bonds and one or more risky financial assets, with a quite diversified portfolio. To help interpretation, in Table 4 segments are ranked in ascending order of product penetration rates, from households owning only bank or postal deposits to households owning more sophisticated financial products.

Table 4 about here

Table 5 lists transition probabilities among segments. A large percentage of households remained in the same segment over the period between consecutive waves as it is indicated by the large percentages on the diagonal of Table 5. The most stable segment is that where households have only bank or postal deposits. The most dynamic segment is that where households invest mainly in state bonds. In general, households who change segment move over the most similar group: dynamics across very different segments is almost negligible.

Table 5 about here

The statistically significant covariates on the probability of the initial state (Table 6) and on transition probabilities (Table 7) are area of the country where the family lives, head of household's age, gender and education.

Table 6 about here

Table 7 about here

Results are quite consistent with previous research on the topic (see, for example, Browning and Luisardi, 1996 and Wärneryd, 1999). Families who live in the North of the county show a higher probability to be in segments 3 (investments in less risky assets), 4 (state bonds) and 5 (investment in more risky assets) and lower in segments 1 (no investments) and 2 (only

[2] Model estimation was performed with Latent Gold 5.0 Syntax Module (Vermunt and Magidson, 2013) and all models were estimated with several sets of starting values in order to avoid local maxima.

deposits), the contrary is true for families in the South; those living in the Centre tend to belong to segment 3 and not to segment 1. For what concerns the gender of the head of the household, our results reflect the different economic conditions of the groups. With a female head of household the probability is higher for the family to belong to segments 1 and 2, lower to belong to segments 3 and 5 that show a higher level of financial investment. The older the head of the household, the more financially active is the family. Heads of household with no education are positively associated to all segments but 5, the contrary is true for heads of household with a primary or a lower-secondary title. When the head of household has the highest level of education, the family is less likely to belong to market segments 1 and 2.

Families living in the North tend to move to segments 4 and 5, the most financially active ones, and not to segment 1; the contrary is true for those living in the South. Families living in the Centre have with lower probability segment 2 as destination state. When the head has primary and lower- secondary education title, households are more likely to move towards segment 5, the contrary is true for households where the head has no education. Highest educated heads show a negative association with market segments 1 and 2 as destination in dynamics over time. When the head of household is female, the probability is higher to move over time to segments 1, 2 and 3, where the level of investments is lower. The pattern of associations among education and the state of destination of the latent chain is very similar to that of the associations with the initial state.

In order to pursue the scope of the paper, a conditional latent class growth model with linear growth and time-varying predictors was specified and estimated on the same dataset. The hypothesis is that the market is segmented and that ownership behavior evolves over time following different patterns in the segments influenced by household's and head's characteristics.

Response variables are, again, the 13 binary indicators y_{ijt} assuming value *1* if household *i* holds financial product *j* at time *t*, *i* varies over *1* to *n*, sample dimension, *j* over 1 to 13 and *t* over 1 to 5, since 5 survey waves are considered, from 2000 to 2010. The same variables describing household's and head of household's characteristics, introduced in the LCM model, were considered as potential covariates for the latent trajectories and the latent classes. The latent trajectory is assumed linear and parameters are assumed constant for all financial products in each segment but different across segments. Potential covariates on the parameters of the trajectory have effects that may differ across segments. With this specification, the latent trajectories summarize the behavior of consumers in each segment of the financial market. Moreover, the model aims at explaining the different latent trajectories by household and head of household characteristics.

The variances of the intercept and the slope are set equal across all classes in order to specify a more parsimonious model and not to increase the computational burden. The model was estimated starting with one latent class and then the number of latent classes was increased till the Bayesian Criterion Index (BIC) increased . The best fitting model was that with four latent classes. Table 8 lists estimation results: segment sizes and response probabilities. Segment 1 has dimension equal to 21.66% and it is that comprising households who hold only deposits. Segment 2 (3.31%) is composed of households that in a non-negligible percentage do not hold a bank or postal deposit, however, they invest in financial assets. The biggest segment has dimension 67.33%: families hold a deposit and moderately invest in other financial assets. Segment 4 (6.72%) is composed of families who are active in the financial market with investments in state bonds and all other types of assets. Segments in the table are ordered from the least to the most financially active.

Table 8 about here

Table 9 lists the results of the best fitting model according to the BIC index. Assuming a linear trajectory over time, the intercept and the slope are significantly different across the segments. The intercept constant term is positive and statistically significant for all segments, decreasing in magnitude from segments 1 to 4. The intercept is significantly affected by the area of the country where the family lives and head of households' gender, age and education. Age of the head of household has a significant and positive effect on the intercept of the trajectory in segment 3 for heads with age between 35 and 49 years old and in segment 1 for the oldest heads, this means that, for example, in segment 3 and for a household with a head who is in the middle age, the slope of the latent trajectory is increased by 0.0876. The average level of financial product ownership behavior significantly increases in moving from the North to the South of Italy in segment 3 and decreases in segment 1. The level of education has a significant effect on the slope only in segment 3, that representing families holding a deposit and making moderate investments in the financial market. Finally, the gender of the head of household has a significant impact on the average level of the behavior only in segment 3, being positive for females. The constant term of the slope of the trajectory is significant and positive for all segments, however, with different magnitudes. The area of the country where the family lives affects positively the slope in segments 1, 3 and 4 for the South, negatively in segments 2, 3 and 4 for the North and in segments 3 and 4 for the Centre. This means, for example, that for families living in the South of the country and belonging to segment 1, the slope of the trajectory is increased by 0.1950. There is a significant effect on the slope by the age of the head of household, this effect has a different pattern in the four segments. Gender of the head of household has a significant and positive impact for females only in segment 2. Age of the head of household has a negative effect in segment 1 for the oldest heads.

Table 9 about here

Table 10 about here

From Table 10, it is possible to see that the variances of the intercept and the slope and the covariance are significantly different from 0. This means that there is substantial variation among households in the initial condition and in the evolution over time within each class.

Summarizing, the two models perform quite differently on our dataset (Table 11). The number and the characteristics of the identified latent states or classes, our market segments, is different. All considered covariates have a significant impact on both models, however, the patterns of the effects appear more sensible in the LCM model on the basis of evidences on the Italian market of financial products reported in the reference literature (see, for example, Bank of Italy, 2012, Guiso *et al*., 1996, Albareto *et al*., 2008). The LCM model has a better fit to the data, however, the LCG model is more parsimonious.

Results of estimation on the dataset used as an example allow us to better understand the specificities of the two approaches.

If the focus of the researcher, or the marketing strategist, is on patterns across segments followed by each customer over time, the LCM model is the most appropriate instrument. With estimated posterior probabilities, individual profiles in the market, i.e., movements across segments over time, may be described and predicted. Observed heterogeneity due to customers' characteristics is also considered. With reference to the example, after estimating the best fitting LCM model, it is possible to forecast the probability to own each financial product at time $t+1$ performing the following steps: first household latent class membership at time $t+1$ is predicted on the bases of latent class membership at time t and in combination with covariates values that have a significant effect on segment belonging and on transition probabilities.

If the focus is on customers' trajectories over time inside each market segment, the LCG model has to be used. In this latter case, segments are identified at the beginning of the observational period and customers are not allowed to change segment over time. However, the conditional approach takes into account customers' characteristics that may influence the estimated trajectory so that observed variability inside each segment is taken into account. After estimating the best LCG model, it is possible to predict the trajectory describing ownership of each financial product for households with the same characteristics on the bases of the significant covariates and in the four market segments.

For what concerns marketing strategy design, the two models provide to the decision maker the same kind of information which consists in being able to identify which type of households inside each market segment can be more prone to acquire other financial products.

Table 11 about here

6. Conclusion

Dynamic market segmentation is a very important topic in many businesses where it is interesting to gain knowledge on the reference market and on its evolution over time. Various papers in the reference literature are devoted to the topic and different statistical models are proposed.

In this paper two statistical approaches to model categorical longitudinal data to perform dynamic segmentation are compared. The LCM model identifies a latent variable whose states represent market segments at an initial point in time, customers can switch to one segment to another between consecutive measurement occasions and a regression structure models the effects of covariates, describing customers' characteristics, on segments belonging and on transition probabilities. The LCG model estimates individual trajectories, describing a behaviour over time, the latent classes identify subgroups with different change patterns. Customers' characteristics may be inserted in the model to affect trajectories and trajectories may vary across latent groups, in our case, market segments.

In our application, customers are households and the market is that of financial products. We refer to financial products ownership by Italian families in the period from 2002 to 2010. In the LCM context, our best fitting model has a stationary first-order latent chain with five latent states, a constant measurement error component and concomitant variables. In The LCG context, the best fitting model is a conditional one with four classes and a linear growth.

The two approaches fit the data very differently with the LCM model showing the best fit in terms of the BIC index. The LCG model, on the other hand, is more parsimonious since it describes with a very simple parametric function, the linear one in our application, customers behaviour over time. The highest level of parsimony is balanced by the assumption that units belong to the same market segment over the observational period, while they are allowed to switch in the LCM model specification.

Estimation results provide substantially the same type of information to the marketing operator both in terms of market knowledge - market segments are identified, and in terms of the capability of predicting ownership behaviour - households typologies in each segment more prone to acquire other products may be detected. Both models identify the same households' characteristics, gender, age and education of the head of household and the area of the country were the family lives as significant in affecting customers' behaviour in the reference market. However, the number of segments and their characteristics is different with the two models.

Moreover, it is important to recognize that the two approaches rely on quite different assumptions: the LCM model allows customers to switch segment over the observational period, while the LCG model does not.

The advice to marketers is to explore both solutions to dynamically segment the reference market. The best approach will be then judged in terms of fit, substantial results and assumptions on the reference market.

The contribution of this paper to the rich exiting literature on the topic of segmentation in the bank and financial services marketing covers various direction. First of all, dynamic segmentation is performed, considering the fact that actual and potential customers may change attitude and behaviour towards the market. Segments are not identified on the bases of demographic characteristics as in also recent studies (Lees *et al.*, 2016) but on the bases of ownership of the different financial products. Socio-demographic characteristics are used to profile segments and eventually explain changes of behaviour over time. The limits of demographic segmentation are already discussed, for example, in Machauer and Morgner (2001) and Pierce *et al.* (2011). Another important improvement regards data quality, since information is collected on a large random sample of Italian households by means of an official survey conducted by the Bank of Italy. In the reference literature, many studies, rely on information collected on non-random convenience samples (Patisotis *et al.*, 2012).

Some further research may regard the specification and the estimation of more complicated models, especially in the LCG approach. The form of the latent trajectory, that might be different from the linear one, may be investigated, and parameters of the latent trajectories may differ across products. All these extensions will imply to face computational and identification problems but they might deserve some attention.

References

Albareto, G., Bronzini, G., Caprara, D., Carmignani, A. and Venturini, A. (2008), "The real and financial wealth of Italian households by region", in Bank of Italy (Ed.) *Household wealth in Italy*, Rome, pp. 57-77.

Bank of Italy (2012), *Supplement to the Statistical Bulletin. Sample Survey Household Income and Wealth in 2010*, Rome.

Bassi, F. (2012), "Analysing markets within the latent class approach: an application to the pharma sector", *Applied Stochastic Models in Business and Industry*, Vol. 29, pp. 199-207.

Connel, A.M. and Frye, A.A. (2006), “Growth mixture modelling in developmental psychology: overview and demonstration of heterogeneity in developmental trajectories of adolescent antisocial behaviour”, *Infantant and Child Development*, Vol. 15, pp. 609-621.
Browning, M. and Lusardi, A. (1996), “Household saving: micro theories and micro facts”, *Jounal of Economic Literature*, Vol. 34: 1797-1855.
Dayton, C.M. and Mcready, G.B. (1988), “Concomitant-variable latent-class models”, *Journal of the American Statistical Association*, Vol. 83, pp. 173-178.
Dempster, A.P., Laird, N.M. and Rubin, D.B. (1977), “Maximum likelihood for incomplete data via the EM algorithm”, *Journal of the Royal Statistical Society, Series B*, Vol. 39, pp. 1-38.
Giraldo, A., Rettore, E. and Trivellato, U. (2001), “Attrition bias in the Bank of Italy’s Survey of Household Income and Wealth”, Working paper 41, Progetto di ricerca “Occupazione e disoccupazione in Italia: misura e analisi dei comportamenti”, Department of Statistical Sciences, University of Padova, Italy.
Guiso, L., Jappelli, T. and Terlizzese, D. (1996), “Income risk, borrowing constraints, and portfolio choices”, *American Economic Review*, Vol. 86 No 1, pp. 158-170
Lees, G., Winchester, M. and De Silva, S. (2016), “Demographic product segmentation in financial services products in Australia and New Zealand”, *Journal of Financial Services Marketing*, On-line first.
Machauer, A. and Morgner, S. (2001), “Segmentation by expected benefits and attitudes”, *International Journal of Bank Marketing*, Vol. 19, pp. 6-18.
Magidson, J. and Vermunt, J.K. (2002), “Latent class models for clustering. A comparison with K-means”, *Canadian Journal of Marketing Research*, Vol. 20, pp. 36-43.
Nagin, D. (2005), *Group-based modeling of development*, Harvard University Press.
Paas, L.J., Bijmolt, T.H.A. and Vermunt, J.K. (2007), “Acquisition patterns of financial products: a longitudinal investigation”, *Journal of Economic Psychology,* Vol. 28 No 2, pp. 229-241.
Patsiotis, A.G., Hughes, T. and Webber, D.J. (2012), “Adopters and non-adopters of internet banking: a segmentation study”, *International Journal of Bank Marketing*, Vol. 30, pp. 20-42.
Pierce, N., Campbell, C. and Heinrich, D. (2011), “Suboptimal segmentation: assessing the use of demographics in financial services advertising”, *Journal of Financial Services Marketing*, Vol. 16, pp. 173-182.
Prinzie, A. and van der Poel, D. (2011), “Modeling complex longitudinal consumer behaviour with dynamic Bayesian networks: an acquisition pattern analysis application”, *Journal of Intelligent Information Systems*, Vol. 36, pp. 283-304.
Rink, D.R., Roden, D.M. and Cox, S.R. (2015), “Risk preferences and the market of financial services: segmentation by birth order”, *Journal of Financial Services Marketing*, Vol. 18, pp. 17-26.
Vermunt, J.K. (2007), “Growth models for categorical response variables: standard, latent-class, and hybrid approaches”, in van Montfort, K., Oud, H. and Satorra, A. (Eds.), *Longitudinal Models in the Behavioural and Related Sciences*, Erlbaum, Mahwah, NJ, pp139-158.
Vermunt, J.K. (2010), “Latent class modelling with covariates: two improved three-step approaches”, *Political Analysis*, Vol. 18, pp. 450-469.
Vermunt, J.K. and Magidson, J. (2013), *LG-Syntax User’s Guide: Manual for Latent GOLD 5 Syntax Module*, Statistical Innovations Inc, Belmont, MA.
Vermunt, J.K., Tran, B. and Magidson, J. (2008), “Latent class models in longitudinal research”, in Menard, S. (Ed.), *Handbook of Longitudinal Research: Design, Meauremenr, and Analysis*, Elsevier, Burlington, MA, pp373-385.

Wärneryd, K.E. (1999), *The Psychology of Savings: a Study of Economic Psychology*, Elgar, Northampton.
Wiggins L.M (1973), *Panel Analysis: Latent Probability for Attitude and Behavior Processes*,Elsevier Scientific, New York.
Zuccaro, C. and Savard, M. (2010) “Hybrid segmentation of internet bank users”, *International Journal of Bank Marketing*, Vol. 28, pp. 448-464.

TABLES

Table 1. Italian Survey on Household Income and Wealth survey plan 1987-2010

First interview	**Survey wave**											
	1987	**1989**	**1991**	**1993**	**1995**	**1998**	**2000**	**2002**	**2004**	**2006**	**2008**	**2010**
1987	8,027	1,206	350	173	126	85	61	44	33	30	28	23
1989		7,068	1,837	877	701	459	343	263	197	159	146	123
1991			6,001	2,420	1,752	1,169	832	613	464	393	347	293
1993				4,619	1,066	583	399	270	199	157	141	124
1995					4,490	373	245	177	117	101	84	75
1998						4,478	1,993	1,224	845	636	538	450
2000							4,128	1,014	667	475	398	330
2002								4,406	1,082	672	525	416
2004									4,408	1,334	995	786
2006										3,811	1,143	856
2008											3,632	1,145
2010												3,330
Total	8,027	8,274	8,188	8,089	8,135	7,147	8,001	8,011	8,012	7,768	7,977	7,951
Panel proportion %	0	14.6	26.7	42.9	44.8	37.3	48.4	45.0	45.0	50.9	54.5	58.1

Table 2. Ownership of financial assets by type. Percentage of households. 2002 -2010

	2002	2004	2006	2008	2010
CD	2.3	2.4	2.6	2.9	3.2
CT	0.9	0.7	1.1	1.7	0.9
BFP	5.9	5.8	6.9	6.5	5.9
BOT	10.1	6.9	8.1	10.2	8.5
BTP	2.7	2.9	2.8	2.6	2.9
OBB	7.7	8.7	8.4	10.4	11.2
QFC	12.6	12.5	12.1	8.7	8.8
GP	2.6	2.0	1.7	0.7	1.0
TE	1.3	1.2	1.1	1.4	1.9
COOP	2.2	2.9	2.5	2.8	2.7
DEP	88.5	89.0	90.7	90.9	85.6
SHA	10.6	8.7	8.8	7.4	6.7
CCT	3.9	3.7	3.8	3.0	2.9

Table 3. Covariates' frequency distribution. Percentage of households. 2002 -2010

	2002	2004	2006	2008	2010
EDUCATION					
No certificate	5.6	4.0	5.2	4.7	4.5
Primary school	28.2	29.2	28.3	28.5	27.9
Lower-secondary	28.1	27.2	27.1	33.0	33.4
Upper-secondary	29.3	29.8	30.2	24.4	25.2
University	8.8	8.9	9.2	9.5	9.1

AGE					
≤ 34	5.6	3.4	2.2	1.2	0.6
35-49	27.6	25.7	22.8	20.3	16.8
50-64	36.3	35.5	35.4	35.6	36.7
≥65	30.5	35.4	39.5	42.9	45.9
GENDER					
Male	68.4	66.1	64.4	63.7	57.1
Female	31.6	33.9	35.6	36.3	42.9
AREA					
North	45.3	45.3	45.3	45.3	45.3
Centre	19.7	19.7	19.7	19.7	19.7
South	35.0	35.0	35.0	35.0	35.0

Table 4. LCM model estimation: segments' sizes and profiles (percentages)*

	1 – no investments	2 – only deposits	3 – less risky assets	4 – state bonds	5 – more risky assets
Size	12.42	50.10	19.57	9.23	8.68
DEP	**75.94**	**99.61**	**99.47**	**99.20**	**100**
BFP	1.10	2.79	**19.28**	3.58	6.65
COOP	0.00	0.12	**9.62**	4.28	3.54
BOT	0.00	1.20	4.89	**64.74**	14.69
BTP	0.00	0.08	0.48	**14.33**	15.22
CCT	0.09	0.08	1.49	**21.37**	13.08
CD	0.28	0.67	6.72	5.56	**5.72**
CT	0.00	0.02	1.95	0.91	**6.44**
OBB	0.00	0.69	20.34	15.13	**40.32**
QFC	0.00	1.14	25.95	9.69	**50.00**
GP	0.00	0.32	3.80	0.31	**7.46**
TE	0.00	0.18	1.89	0.55	**9.75**
SHA	0.00	0.77	9.85	3.75	**64.90**

* To help interpretation, some meaningful percentages appear in bold

Table 5. LCM model estimation: transition probabilities

	1 – no investments	2 – only deposit	3 – less risky assets	4 – state bonds	5 – more risky assets
1- no investments	0.7964	0.2036	0.0000	0.0000	0.0000
2 – only deposit	0.0288	0.9365	0.0101	0.0209	0.0037
3 – less risky assets	0.0000	0.0604	0.9068	0.0328	0.0000
4 – state bonds	0.0000	0.0790	0.1182	0.7836	0.0192
5 – more risky assets	0.0000	0.0729	0.0292	0.0247	0.8732

Table 6. LCM model estimation: covariates' effects on initial state (effect coding)

	1 – no investments	2 – only deposit	3 – less risky assets	4 – state bonds	5 – more risky assets
GENDER					
Female	0.3761*	0.1548*	-0.3140*	0.1100	-0.3268*

AGE					
≤ 34	1.0326*	0.4652*	0.2276	-1.1674	-0.5581
35-49	-0.3433	0.2265*	0.2168	-0.2300	0.1300
50-64	-0.4633*	-0.2813*	-0.1186	0.6276*	0.2386
≥65	-0.223o	-0.4105*	0.4391*	0.7698*	0.1895
EDUCATION					
No title	44.2614*	42.6165*	19.4114*	42.0793*	-148.3686*
Primary school	-8.9257*	-10.0561*	-5.1236*	-10.5902*	34.6956*
Lower-secondary	-10.3643*	-10.6875*	-11.5824*	-10.7931*	36.9217*
Upper-secondary	-11.5824*	-10.9163*	-4.7990	-10.5196	37.8174
University	-13.3890*	-10.9566*	-4.4120	-10.1763	38.9340
AREA					
North	-1.2565*	-0.4486*	0.4361*	0.3581*	0.9109*
Centre	-0.7094*	-0.1450	0.5644*	0.1373	0.1526
South	1.9659*	0.5936*	-1.0005*	-0.4954*	-1.0635*

* significant at 0.05

Table 7. LCM model estimation: covariates' effects on state at time t+1 (effect coding)

	1 – no investments	2 – only deposit	3 – less risky assets	4 – state bonds	5 – more risky - assets
GENDER					
Female	0.2164*	0.0087*	0.3732*	0.1022	-0.7005*
AGE	0.0346	-0.873*	-0.5510	-1.0982	-0.9196
≤ 34	-0.7259*	-0.9117*	-1.1897*	-0.6599*	3.4871*
35-49	1.5329*	1.2144*	-0.1718	-0.8193*	3.7383
50-64	-2.7151*	-0.8778*	-0.7173	-0.1156	4.4258
≥65	-0.3087	-0.2679	-0.1304	0.4401*	0.2669
EDCATION					
No title	4.7812*	3.7531*	2.7939*	2.3766*	-13.7048*
Primary school	0.2122	-0.6662*	-0.8702*	-1.0563*	2.3804*
Lower-secondary	-0.5210	-0.9130*	-0.9685*	-0.7558*	3.1582*
Upper-secondary	-1.4512*	-1.2305*	-0.4915	-.05706	3.7438*
University	-3.0213*	-0.9434*	-0.4637	0.0061	4.4223*
AREA					
North	-0.7117*	-0.1348	-0.4294	0.4905*	0.7854*
Centre	-0.2272	-0.2071*	0.3000	0.0003	0.1340
South	0.9389*	0.3419*	0.1294	-0.4908*	-0.9194*

* significant at 0.05

Table 8. Conditional LCG model estimation: segments' sizes and profiles (percentages)

	1 - only deposits	2 – no deposits and moderate investments	3 deposits and moderate investments	4 – state bonds and risky assets
Size	21.66	3.31	67.33	7.69
DEP	91.68	58.05	93.42	62.09

BFP	0.60	5.42	7.90	8.92
COOP	0.14	2.64	3.08	4.63
TE	0.05	1.54	1.49	2.85
BOT	0.99	6.74	10.46	10.91
BTP	0.13	2.55	2.94	4.48
CCT	0.19	3.06	3.74	5.28
SHA	0.77	6.03	9.06	9.84
CT	0.04	1.28	1.16	2.42
CD	0.15	2.78	3.29	4.84
OBB	0.95	6.60	10.18	10.69
QFC	1.32	7.56	12.12	12.12
GP	0.07	1.88	1.95	3.40

Table 9. Conditional LCG model estimation: covariates' effects on intercept and slope

	1 - only deposits	2 – no deposits and some investment	3 – deposit and some investment	4 – state bonds and risky assets
Intercept				
Constant	3.7113*	2.3577*	1.8383*	1.5020*
GENDER				
Female	-0.0736	-0.1696	0.0680*	0.0681
AGE				
≤ 34	-0.3909	0.0597	-0.0428	0.1299
35-49	0.0144	-0.2187	0.0876*	-0.1226
50-64	0.0543	0.0662	-0.0155	-0.2641
≥65	0.3262*	0.0927	-0.0603	0.2568
EDUCATION				
No title	-1.2461	-0.7683	0.1515	-0.0136
Primary school	-0.9730	-0.7450	0.1534*	0.1693
Lower-secondary	-0.6433	-0.4541	0.1050*	-0.1872
Upper-secondary	0.7751	-0.9799	-0.1310*	0.3250*
University	2.0873	2.9472	-0.2789*	-0.2935
AREA				
North	0.3800	-0.0829	-0.1966*	-0.0577
Centre	0.6432	0.0116	-0.1402*	0.0781
South	-1.0232*	0.0714	0.3367*	-0.0204
Slope				
Constant	0.2797*	1.1251*	0.0555*	0.9566*
GENDER				
Female	0.0248	0.1713*	-0.0002	-0.0174
AGE				
≤ 34	0.1942	-0.0235	0.0674	-0.1610
35-49	0.0088	-0.1395	-0.0156	0.0218
50-64	-0.0589	0.1124	-0.0282	0.0729
≥65	-0.1444*	0.0506	-0.0236	0.0662
EDUCATION				
No title	0.2797*	1.1251*	0.1577*	0.5249*

Primary school	0.1769*	2.3467*	0.0149	0.1458
Lower-secondary	-0.0524	-0.5193	-0.0400*	-0.2000*
Upper-secondary	-0.2022*	-1.0664*	-0.0515*	-0.2300*
University	-0.2684	-0.9482	-0.0810*	-0.2408*
AREA				
North	-0.1242	-0.1211*	-0.0465*	-0.8869*
Centre	-0.0708	-0.0121	-0.0194*	-0.7914*
South	0.1950*	0.1332	0.0659*	1.6783*

* significant at 0.05

Table 10. Variances and covariances

	Coefficient	Standard error
Ψ_{α}	0.1269*	0.0154
Ψ_{β}	0.0165*	0.0021
$\Psi_{\alpha\beta}$	0.0021*	3.6434

* significant at 0.05

Table 11. Models comparison

	npar	BIC	L^2	# latent classes
LCM	169	37,110	35,408	5
LCG	107	40,575	39,399	4