Available online at www.sciencedirect.com
SCIENCE DIRECT®
European Journal of Operational Research 170 (2006) 710–734
EUROPEAN JOURNAL OF OPERATIONAL RESEARCH
www.elsevier.com/locate/ejor

Interfaces with Other Disciplines

Investigating purchasing-sequence patterns for financial services using Markov, MTD and MTDg models

Anita Prinzie, Dirk Van den Poel *

Department of Marketing, Faculty of Economics and Business Administration, Ghent University, Hoveniersberg 24, B-9000 Ghent, Belgium

Received 1 August 2003; accepted 7 May 2004
Available online 27 July 2004

Abstract

In the past, several authors have found evidence for the existence of a priority pattern of acquisition for durable goods, as well as for financial services. Its usefulness lies in the fact that if the position of a particular customer in this acquisition sequence is known, one can predict what service will be acquired next by that customer. In this paper, we analyse purchase sequences of financial services to identify cross-buying patterns, which might be used to discover cross-selling opportunities as part of customer relationship management (CRM). Hereby, special attention is paid to transitions, which might encourage bank-only or insurance-only customers to become financial-services customers. We introduce the mixture transition distribution (MTD) model as a parsimonious alternative to the Markov model for use in the analysis of marketing problems. An interesting extension on the MTD model is the MTDg model, which is able to represent situations where the relationship between each lag and the current state differs. We illustrate the MTD and MTDg model on acquisition sequences of customers of a major financial-services company and compare the fit of these models with that of the corresponding Markov model. Our results are in favor of the MTD and MTDg models. Therefore, the MTD as well as the MTDg transition matrices are investigated to reveal cross-buying patterns. The results are valuable to product managers as they clarify the customer flows among product groups. In some cases, the lag-specific transition matrices of the MTDg model give better insight into the acquisition patterns than the general transition matrix of the MTD model.

Keywords: Banking; Marketing; Mixture transition distribution (MTD) models; MTDg models; Cross-sell; Financial-services industry; Analytical customer relationship management (CRM); Business intelligence; Data mining

* Corresponding author. Tel.: +32 9 264 89 80; fax: +32 9 264 42 79.
E-mail address: dirk.vandenpoel@ugent.be (D. Van den Poel).

0377-2217/$ - see front matter
doi:10.1016/j.ejor.2004.05.004

1. Introduction

Sequences have been the subject of research in many disciplines, among which archaeology (McBrearty, 1988), biology (e.g., DNA sequence analysis–Raftery and Tavaré, 1994; Lipman and Pearson, 1984; or even bird songs–Chatfield and Lemon, 1970), chemistry (Xu and Agrawal, 1996), computer sciences (Sabherwal and Robey, 1995), economics (Hopp, 1987), econometrics (Bollerslev et al., 1992), history (Abbott, 1995), linguistics (Jonz, 1989), meteorology (Raftery, 1985a; MacDonald and Zucchini, 1997), psychology (Cohen et al., 1990) and sociology (Abbott and Hrycak, 1990; Katz and Proctor, 1959; Logan, 1981). We define *a sequence* as a succession of events. An *event* is a transition from one discrete state to another, situated along a time continuum (Abbott, 1995). On the one hand, this time aspect can be used for identifying the order of events. On the other hand, 'time' could give precise information on when a certain event happened. The purpose of sequence analysis can be threefold: (1) a general description of sequences, (2) the revelation of the determinants of the transitions and (3) the prediction of the transitions.

In marketing, sequence analysis is applied in choice modelling and is mainly focused on the succession of purchases. Consumer behaviour can indeed be seen as a succession of events (e.g. decisions/choices). The rationale behind the existence of typical purchase sequences is twofold: (1) logical succession due to complementary goods (Feick, 1987), e.g., a VCR is bought after a television set is acquired, and (2) utility maximisation under budget constraints (Hauser and Urban, 1986).

Several authors have shown that a *typical* acquisition pattern exists (i.e., priority pattern) for consumer durables (Pyatt, 1964; Paroush, 1965; McFall, 1969; Hebden and Pickering, 1974; Lush et al., 1978; Kasulis et al., 1979; Clarke and Soutar, 1982; Mayo and Qualls, 1987). In Table 1, we provide an overview of some important characteristics of these studies. Most of them, except for the study by Mayo and Qualls (1987), use cross-sectional data to identify the priority pattern. They assume that if the sample represents a true cross-section of the population, individuals at all stages in the order of the acquisition process should be present in the sample. Except for the study by McFall (1969), the priority patterns are based on reported ownership data. McFall argues that past ownership data reflect past purchase priority patterns. However, they do not provide, except by inference, the current priority patterns maybe governing tomorrow's buying decisions. For the latter purpose, purchase intentions should be used. Furthermore, acquisition patterns have been observed for both homogeneous and heterogeneous sets of products. A homogeneous set of products consists of goods capable of performing similar or related functions, whereas a heterogeneous set of goods is able to perform quite different and unrelated functions. The choice between the two types of sets is mainly based on whether the author believes that households make a priority list between products fulfilling similar purposes or whether also products different in nature are competing for the scarce resources. All studies show evidence for the existence of a universal priority pattern of consumer durables. However, small deviations from this unique priority pattern are possible due to differences in social class, dwelling unit, income, family life cycle and more generally culture (Clarke and Soutar, 1982). Finally, most studies are descriptive and explanatory in nature giving a general description of the acquisition sequence and explaining the determinants of the transitions in the priority pattern. Only the study of Mayo and Qualls (1987) also has a predictive perspective. It investigates if the discovered acquisition pattern is a good predictor of the real acquisition behaviour.

Contrary to the substantial amount of research studying the acquisition sequence for consumer durable goods, research on acquisition patterns in the context of financial assets has been more limited. After having analysed priority patterns for consumer durables, Stafford et al. (1982) also found evidence for the existence of a common acquisition pattern for financial services. They observe a sequence from checking/savings accounts to insurance to stocks, bonds and mutual funds. The acquisition sequence is relatively constant over the three cohort groups defined, except for the purchase of stocks and bonds. Besides this descriptive study, Kamakura et al. (1991) presented a descriptive as well as predictive study. They investigated the

Table 1
Literature review of research on acquisition patterns of household durables

Reference	Type of data	Sample	Data collection method	Type of set of durables	Method	Pattern based on	Covariates	Purpose
Pyatt (1964)	Cross-sectional	Britain		Homogeneous set of electrical appliances	Dynamic Bayesian differential equation model		Social class (1)	1&2
Paroush (1965)	Cross-sectional	(1) 5000 Jewish families (Israël) (2) 1300 families (Israël) (3) 2600 British families	(1) Survey (2) Survey (3) Survey	Homogeneous set of 4 household durables	-Guttman-scalogram analysis -Point correlation matrix	Reported ownership data	-Continent of birth (0) -Duration of residence (0)	1&2
McFall (1969)	(1) Cross-sectional for discovering past purchase priorities (2) Longitudinal to predict future purchase priorities	United States: 943 subscribers to Consumers Union in Midwest	(1) Survey (2) Two interviews separated by a 5 month period	Homogeneous set of durables	Guttman-scalogram analysis	(1) Reported ownership data (2) Consumer purchase intentions	-Urbanisation: urban vs rural (0) -Income: upper or lower (1)	1&2 (& 3)
Hebden and Pickering (1974)	Cross-sectional	386 Heads of households in 50 different areas of Britain	Personal interview and survey	4 Homogeneous sets and 1 heterogeneous set	Matrix of conditional probabilities (extension on method used by Pyatt)	Reported ownership data	-Social class (1, for five leisure goods) -Family lifecycle (1, for five diverse goods)	1&2

Lush et al. (1978)	Cross-sectional	USA: 1800 households	Survey	Homogeneous set of household durables	A multiple criteria evaluation procedure	Reported ownership data		1
Kasulis et al. (1979)	Cross-sectional	Oklahoma City Standard Metropolitan Statistical Area	Personal interview	Heterogeneous set of 12 household durables	Guttman-scalogram analysis	Reported ownership data	-Rent/own dwelling unit (1) -Rent house/ rent duplex/ apartment (1)	1&2
Clarke and Soutar (1982)	Cross-sectional	Australia: 394 heads of households within the Perth metropolitan area	Self-administered questionnaires	Heterogeneous set of 15 household durables	Guttman-scalogram analysis	Reported ownership data	-Rent/own dwelling unit (1)	1&2
Mayo and Qualls (1987)	Longitudinal panel data	USA: 311 household couples married in Illinois during Summer 1968	18 Interviews. Waves are aggregated over time to reflect the stages of the household life cycle	Heterogeneous set of 10 household durables	-Latent structure analysis -Partial least squares	Reported ownership data	-Life cycle (1)	1&2&3

Notes. Column covariates: 1 indicates that the covariate influences the acquisition pattern, 0 indicates that the covariate has no effect on the acquisition pattern. Column purpose: The purpose of these studies can be threefold: (1) a general description of sequences, (2) the revelation of the determinants of the transitions and (3) the prediction of the transitions.

influence of the financial maturity of the customer (cf. life stage) and the acquisition difficulty of the service (cf. resources required, level of risk and liquidity, information costs) on the acquisition sequence of financial services. This way services and customers are positioned along a common 'latent' difficulty/ability continuum expressing the hierarchy of investment objectives (cf. latent-trait analysis). The probability that an investor owns a particular financial service is a function of the investor's position on the continuum relatively to that of the service. The authors hypothesise that more difficult services are acquired in the later stages of the family life cycle. Using the position parameters of the services, Kamakura et al. (1991) discover the same order of acquisition of financial services as suggested in the pyramid of financial independence, which financial advisors use to outline how households should channel their resources into a variety of investments over time. Entirely like Maslow's pyramid of needs, a household moves from the base of this pyramid to its apex corresponding with a hierarchical mode of financial need-satisfaction. Finally, the authors predict the likelihood an investor, given his financial maturity, would acquire each of the financial services, given their acquisition difficulty, with the aim to identify cross-selling prospects. Kamakura et al. (1991) are the first authors using the discovered acquisition pattern for cross-sell purposes. After all, if a unique acquisition pattern for all customers exists, one could perfectly predict what service will be acquired next given the customers' position in the sequence.

Cross-selling pertains to efforts to augment the number of services that a customer uses within a firm. Given increasing acquisition costs, marketers realise that current customers are by far the best prospects for the sales of current and new services. After all, it is easier for institutions to grow by cross-selling services to existing customers than by attracting new ones (Felvey, 1982). Moreover, selling additional products to the same customer has an important positive influence on the relationship between buyer and seller (Van den Poel and Larivière, 2004). This results mostly in an increasing total value of the customer over his lifetime (Kamakura et al., 2003) and a decreasing chance for the customer to end the relationship. The positive effect on retention results from increasing switching costs with multiple relationships (Srivastava and Shocker, 1987) and of the increased ability of the firm to satisfy the customer's needs more effectively than competitors (Kamakura et al., 2003).

Despite the importance of cross-selling for a company's profitability, many enterprises perform their cross-sell actions based on intuition and experience. Also surprisingly, cross-selling received limited attention in the academic literature. The study of Kamakura et al. (1991) requires that the firm knows about each customers' usage of services from both the firm and its competitors, which is unlikely. To accommodate these requirements, Kamakura et al. (2003) propose a mixed factor data analyser. The latter combines information from a survey on the use of financial services outside the firm with data from the customer database on service usage within the firm, to predict ownership of services with the financial-services provider and its competitors. These predictions support the selection of cross-selling prospects.

In this paper, we analyse acquisition sequences for financial services. If sequential patterns emerge from this descriptive study, this may lay the foundation of a subsequent predictive analysis where acquisition prediction rules are deduced from the observed acquisition patterns on the training sample and further applied on a validation sample to identify cross-sell possibilities. Over the last two decades, the financial-services market has become extremely competitive due to increasing consolidation into fewer but larger players on the one hand and deregulation on the other hand, resulting in diminishing profit margins and blurring distinctions between banks, insurers and brokerage firms (cf. universal banking, Berger et al., 1999; Berger, 2003). Hence, nowadays a small number of large institutions offering a wider set of services, dominate the financial-services industry. These developments stimulated bank assurance companies to shift away from a product focus to a customer-oriented approach leading to the implementation of customer relationship management (CRM). An important aspect of this customer orientation is the use of customer transaction databases for cross-selling services (Larivière and Van den Poel, 2004). In a highly competitive environment, the general advantages of cross-selling exceed the possible gains from new customer acquisition, as the acquisition of new customers mostly happens at the expense of competitors and as these new customers tend to be switchers (Kamakura et al., 2003; Buckinx and Van den Poel, forthcoming). Since most financial-services groups have their origin in several mergers between bank and insurance companies, cross-selling insurance products to bank-only customers (i.e., customers who only possess bank products) and cross-selling bank products to insurance-only customers (i.e., customers who only acquired insurance policies) are a major concern. An in-depth analysis of the acquisition sequences is the foundation of an efficient cross-sell strategy. In the past, purchase sequences have been described as: (1) a hierarchical process (cf. Guttman-scalogram analysis and latent-trait analysis), (2) a succession of purchases (cf. Markov models, sequential association rules and optimal matching), or (3) a sequence with the focus on the time aspect (cf. survival analysis). In this study we opt for the second approach. Contrary to hierarchical models, sequence models do not assume the acquisition patterns to be hierarchical. A possible purchase pattern might be customers buying a television set and later on a VCR. Where hierarchical models suppose the purchases to be consecutive, this is not an assumption of sequence models (Agrawal and Srikant, 1995).

The different sequence analysis techniques can be classified by the length of the considered succession of events (Abbott, 1995). At first, scientists focused on methods considering only one transition (first-order dependence). Markov models are the most popular methods. They calculate transition probabilities based on the transition between two events (e.g. purchase events). Gradually, higher-order dependencies were taken into account. This gave rise to *n*th order Markov models and also to 'sequential association rules' and the 'optimal matching' method. Whereas Markov models treat sequences step-by-step (cf. transitions from one state to another discarding the sequence as a whole), the latter methods treat them as whole units. The central issue is whether there are patterns in the sequences, either over the whole sequences or within parts of them (Srikant and Agrawal, 1996). The optimal matching method calculates a similarity measure between the sequences, which is further processed by clustering algorithms to typical sequential patterns (Abbott and Hrycak, 1990).

In this paper we apply *n*th order Markov models as well as a recently introduced technique, the mixture transition distribution model (MTD) (Raftery, 1985a). The MTD model allows estimating high-order Markov chains as it is far more parsimonious than the fully parameterised Markov model. Moreover, the MTD model provides a smaller transition matrix facilitating managerial interpretation. Finally, we also test the performance of the MTDg model, which is less parsimonious than the MTD model, but is capable of representing situations where the relationship between each lag (for instance $t-1$ is the first lag) and the period t is different. It provides lag-specific transition matrices giving additional insight into the acquisition process.

This paper contributes to existing research in three ways: *Methodologically*, we compare different techniques (Markov, MTD and MTDg) for the description of sequences. We provide a parsimonious model to estimate high-order Markov chains, while still being able to interpret the results easily and translating them into managerial conclusions. *Methodologically for marketing*, we demonstrate that the parsimony of the MTD model is also welcome in case of a random variable with a high number of possible values (typically in a marketing context) and not necessarily a high-order dependence. Moreover, we illustrate how the small size transition matrices from the MTD and MTDg models can easily be translated into flow charts, giving a quick insight into the acquisition patterns present. *Empirically for marketing*, we describe the acquisition sequences of financial services for customers having at least three previous purchase events, which might contribute to a consecutive predictive study with the aim to identify cross-sell possibilities. Unlike the studies of Kamakura et al. (1991, 2003), we identify possible paths to convert bank-only or insurance-only customers into financial-services customers, a hot issue in the financial-services market.

The remaining parts of this paper are structured as follows: In the next section we describe the characteristics of the Markov, MTD and MTDg model. We illustrate these models on a financial-services application. Subsequently, we discuss the observed acquisition patterns of the best model from a cross-buying perspective. Here, the appealing interpretational features of the MTD and MTDg model are revealed. Several managerial conclusions are drawn. Finally, we round up this paper with a discussion of the limitations and suggestions for further research.

2. Methodology

2.1. Introduction

The MTD model was introduced by Raftery (1985a) as a parsimonious model for high-order Markov chains with a finite state space. Although Markov chains are well suited to represent high-order dependencies between successive observations of a random variable, they tend to result in a large number of parameters to estimate. As the order k of the chain and the number m of possible values of the random variable X increase, the number of independent parameters increases exponentially and becomes rapidly too large to estimate efficiently or even identifiably (Berchtold and Raftery, 2002). On the other hand, the MTD model involves only one additional parameter for each extra lag (Raftery, 1985a).

Besides being more parsimonious, the MTD model is also attractive from a managerial perspective. Firstly, whereas the transition matrix of a high-order Markov model is hard to interpret managerially, the MTD model overcomes this caveat by giving a short form $m \times m$ size transition matrix capturing the overall tendencies. Secondly, the MTD model provides lambda-weights representing the importance of each lag on the current state.

The following sections are structured as follows: In Section 2.2 we review the concept of Markov chains and we point out the relevance of parsimonious models for high-order Markov chains. In the next Section, we define the MTD model. Subsequently, we elaborate on one of the MTD model generalisations, the MTDg model. In Section 2.5, we discuss how the models are evaluated and compared. Section 2 is essentially based on Raftery (1985a), Raftery and Tavaré (1994), Berchtold (2001), and Berchtold and Raftery (2002).

2.2. Markov chains

The Markov chain, introduced by Markov at the beginning of the 20th century, is a probabilistic model used to represent dependencies between successive observations of a random variable. It is applied in many domains including meteorology, geography, biology, chemistry, physics, behavioural and social sciences.

In this paper, we consider a discrete-time random variable X_t taking on values from the finite set $N = \{1, \ldots, m\}$. Our goal is to describe the value taken by X_t as a function of the values taken by k previous observations of this same variable. For example, let X_t be the monthly purchase from a direct-mail company in one of the following three product categories: (1) books, (2) CDs, (3) DVDs. Hence, X_t takes values in the set $\{1, 2, 3\}$. We want to describe the purchase in one of these categories at moment t as a function of the previous purchases.

In a *first-order* Markov model ($k = 1$) the current value of X_t is fully explained by the value taken by that same variable at time $t-1$ (i.e., the first lag). Hence, we can write

$$P(X_t = i_0 | X_0 = i_t, \ldots, X_{t-1} = i_1) = P(X_t = i_0 | X_{t-1} = i_1) = q_{i_1 i_0}(t), \tag{1}$$

where $i_t, \ldots, i_0 \in \{1, \ldots, m\}$. We assume the probability $q_{i_1 i_0}(t)$ to be time-invariant: $q_{i_1 i_0}$. The result is a *homogeneous* Markov chain.

The model is summarised in a transition matrix Q giving the probability distribution of X_t given any possible value of X_{t-1}. Each row of Q is a probability distribution (i.e., each row sums to one and the elements are nonnegative).

$$Q = \begin{array}{c|cccc} & & X_t & & \\ X_{t-1} & 1 & \cdots & \cdots & m \\ 1 & q_{11} & \cdots & \cdots & q_{1m} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ m & q_{m1} & \cdots & \cdots & q_{mm} \end{array}. \tag{2}$$

In *a* kth order Markov chain, the present depends on k previous periods. For example, in a second-order model the current value is explained by all lags up to $t-2$ (i.e., lags 1 and 2). The transition probabilities are

$$P(X_t = i_0 \ |X_0 = i_t, \ldots, X_{t-1} = i_1) = P(X_t = i_0 | X_{t-k} = i_k, \ldots, X_{t-1} = i_1) = q_{i_k \ldots i_0}. \tag{3}$$

The transition matrix of a kth order Markov model is of a larger size. In our direct-mail company example, we assume the discrete-time window to be one month. Given $m=3$ and $k=2$ (i.e., taking the purchases of the last two months into account), the corresponding reduced or collapsed transition matrix R is (Pegram, 1980)[1]

$$R = \begin{array}{cc|ccc} & & & X_t & \\ X_{t-2} & X_{t-1} & 1 & 2 & 3 \\ 1 & 1 & q_{111} & q_{112} & q_{113} \\ 2 & 1 & q_{211} & q_{212} & q_{213} \\ 3 & 1 & q_{311} & q_{312} & q_{313} \\ 1 & 2 & q_{121} & q_{122} & q_{123} \\ 2 & 2 & q_{221} & q_{222} & q_{223} \\ 3 & 2 & q_{321} & q_{322} & q_{323} \\ 1 & 3 & q_{131} & q_{132} & q_{133} \\ 2 & 3 & q_{231} & q_{232} & q_{233} \\ 3 & 3 & q_{331} & q_{332} & q_{333} \end{array}. \tag{4}$$

Each possible combination of k successive observations of the random variable X is a *state*. The number of states is equal to m^k (in example: 3^2=9, see dashed box). In each row of the matrix R, there are $(m-1)$ independent probabilities, as the last transition probability equals one minus the other transition probabilities in that row. The total number of independent parameters to estimate is thus equal to $m^k(m-1)$. From the latter formula, it is obvious that each additional lag involves a substantial increase in number of parameters to estimate, which could prevent the modelling of high-order Markov chains, although high-order dependencies might be present in the data. This is exactly where the MTD model comes into the picture. Finally, the $m^k(m-1)$ parameters are estimated by maximum likelihood procedure (Berchtold and Raftery, 2002).

2.3. MTD model

The MTD model was introduced to approximate high-order Markov chains with far less parameters than the fully parameterised model. Each element of the transition matrix is the probability of observing an event at time t given the events observed at times $(t-k)$ to $(t-1)$. In the MTD model, the effect of each

[1] The reduced transition matrix R is equal to the transition matrix Q without *structural zeros*, i.e., elements corresponding to transitions that cannot occur.

lag upon the present is considered *separately*. As such, it is a simplification of the Markov model, because interaction effects between lags are not considered. The MTD conditional probabilities are a *mixture* of linear combinations of contributions in the past:

$$P(X_t = i_0 | X_{t-k} = i_k, \ldots, X_{t-1} = i_1) = \sum_{g=1}^{k} \lambda_g q_{i_g i_0}. \tag{5}$$

The $q_{i_g i_0}$ are the probabilities of an $m\times m$ transition matrix Q. Compared to the transition matrix resulting from a traditional Markov model with size $m^k\times m$, the MTD model enables to gain insight into the stochastic process, even when high-order dependencies are present. Moreover, interpretation is facilitated as the MTD model provides lambda-parameters $\lambda=\lambda_1,\ldots,\lambda_k$, which are a kind of weight parameters expressing the effect of each lag g on the present value of X (i.e., i_0). In order to ensure the results of the model to be probabilities, the lag parameters are subject to following constraints:

$$\sum_{g=1}^{k} \lambda_g = 1, \tag{6}$$

$$\lambda_g \geqslant 0. \tag{7}$$

Let us compare the definition of the conditional probabilities for Markov and MTD using the direct-mail company example, $m=3$ and $k=2$. Suppose we want to know the probability of buying a book now, given that the customer bought a CD two months ago and a book during the previous purchase occasion.

$$P(X_3 = 1 | X_1 = 2, X_2 = 1).$$

For Markov

$$= P(X_3 = 1 | X_1 = 2, X_2 = 1) = q_{i_{211}}. \tag{8}$$

For MTD

$$= \lambda_1 \times P(X_3 = 1 | X_2 = 1) + \lambda_2 \times P(X_3 = 1 | X_1 = 2) = \lambda_1 q_{i_{.11}} + \lambda_2 q_{i_{2.1}}. \tag{9}$$

The parameters λ and q of the MTD model (5) are estimated using maximum likelihood (Berchtold and Raftery, 2002).

The MTD model is far more parsimonious than the whole parameterised Markov model. The transition matrix Q has $m\times(m-1)$ independent parameters (cf. m columns in Q and $m-1$ independent parameters in each row of Q). In addition, a kth order model has $(k-1)$ independent parameters (cf. lambda's are summing to one). Hence, a kth order MTD model has only $[m\times(m-1)]+(k-1)$ parameters to estimate, which is a lot less, given $k>1$, than the $m^k(m-1)$ parameters of the kth order Markov model. Moreover, notice from the formula that MTD requires only one extra parameter for each supplementary lag (cf. Table 2).

The MTD model can also be used to improve the estimation of a Markov chain, where some rows in the transition matrix R are very poorly estimated due to too few observations (Berchtold and Raftery, 2002). The rows estimated with a large number of data points have almost identical transition probabilities in the MTD matrix, whereas rows estimated with a very small number of observations tend to be smoother in the MTD transition matrix.

Concluding, the MTD model enables us to parsimoniously estimate high-order transition matrices even with a relatively small number of observations and it facilitates the interpretation because of the small size of Q (cf. $m\times m$ for MTD vs $m^k\times m$ for Markov). Moreover, the lambda-parameters give us additional insight into the stochastic process.

Table 2
Maximum number of independent parameters for kth order Markov, MTD and MTDg models for the book and CD retailer example

Number of values, m	Order, k	Markov chain, $m^k \times (m-1)$	MTD model, $[m \times (m-1)]+(k-1)$	MTDg model, $[k \times m \times (m-1)]+(k-1)$
3	1	6	6	6
	2	18	7	13
	3	54	8	20

Remark: The actual number of independent parameters can be lower than the maximum number, because parameters estimated to be zero are not taken into account following the convention in counting degrees of freedom for models for categorical data.

2.4. Extension of the MTD model: The MTDg model

Since the introduction of the MTD model for the modelling of time-homogeneous high-order models with finite state spaces by Raftery (1985a), it has been generalised and extended in many ways. The MTD model has been developed for infinite lags (Le et al., 1996), for time series with missing data and for Markov chains with an infinite denumberable state space (Raftery, 1985b) or with an arbitrary state space, i.e., the GMTD model (Martin and Raftery, 1987; Adke and Deshmukh, 1988; Raftery, 1993; Le et al., 1996; Wong and Li, 2000). The MTD model has even been applied in a spatial context (Raftery and Banfield, 1991; Berchtold, 2001). Another interesting development of the MTD model consists in the use of a different transition matrix for each lag (Raftery, 1985a). This model is called the MTDg model and is the subject of the discussion in this section. Whereas in the original MTD model the same transition matrix Q is used to represent the relationship between each lag and the present, the MTDg model relaxes this assumption by using a different $m \times m$ transition matrix for each lag. These k different transition matrices might deliver additional insight in the stochastic process, especially if there is a completely different relation between each lag and the period t. The MTDg model is defined as

$$P(X_t = i_0 | X_{t-k} = i_k, \ldots, X_{t-1} = i_1) = \sum_{g=1}^{k} \lambda_g q_{i_g i_0}^{(g)}, \tag{10}$$

where $q_{i_g i_0}^{(g)}$ is the probability to observe $X_t = i_0$ and $X_{t-g} = i_g$. As there are k different transition matrices, the MTDg model has $[k \times m \times (m-1)]+(k-1)$ independent parameters. Although the MTDg model is less parsimonious than the MTD model, it still has less parameters than the full Markov model (see Table 2). Moreover, it has the added potential of encouraging the understanding of the stochastic process in case of very different relationships between each lag and the present state. The estimation of the parameters $\lambda_k, \ldots, \lambda_g$ and $Q_1, \ldots, Q_g$ is similar to the maximum likelihood estimation of the parameters in the MTD model, except that k transition matrices have to be estimated instead of just one.

2.5. Model fit and evaluation

To evaluate the fit of each model and to compare different models (Markov, MTD and MTDg), we use the Bayesian Information Criterion (BIC) (Schwarz, 1978). The BIC balances the desire for a better fitting model against the desire for a parsimonious model:

$$\text{BIC} = -2\text{LL} + p \log(n), \tag{11}$$

where LL is the log-likelihood of the model, p is the number of independent parameters and n is the number of components in the log-likelihood. We do not take into account the parameters estimated to be zero, as is indicated by the derivation of the BIC approximation and it is also in line with the convention in counting the degrees of freedom for models for categorical data. The model with the lowest BIC is chosen, which is approximately the same as choosing the model with the highest posterior probability (Raftery, 1985a).

We prefer the BIC to AIC (i.e., Akaike's Information Criterion) statistic, because the former is a more consistent estimator of the true order of the Markov chain (Katz, 1981). Whereas the AIC tends to underestimate the true order of the Markov chain, no matter the sample size, the BIC tends to overestimate the order in case of rather small samples.

3. Financial-services application

The Markov, MTD and MTDg models described above are applied to a database of a major international financial-services provider, referred to as the IFSP, in order to gain insight into the acquisition sequence of financial services. Like many of its competitors, the IFSP went through a wave of mergers between bank and insurance companies resulting in a financial-services group. Nowadays, the IFSP is serving about 50 million individuals and companies worldwide. As the financial-services market has become extremely competitive with diminishing profit margins, knowing the customer is more crucial than ever and may result in competitive advantage. Hence, acquiring insight into the financial services acquisition pattern might be a first step in getting to know the customer.

In this paper, we investigate how customers, having *at least four* distinct purchase moments, acquire financial services and to what extent cross-buying occurs. We aim to find acquisition patterns that might be identifiable as cross-selling opportunities. However, the latter requires a predictive study translating the acquisition patterns discovered on the sample into rules predicting in what product group[2] a customer will buy next. The predictive accuracy of the application of these rules on a test sample then indicates whether these acquisition patterns are a solid basis to predict cross-buying behaviour and, hence, to support cross-sell actions. We focus on customers for whom we observe at least four purchase events. This selection condition seems acceptable, as we wish to demonstrate the parsimony of the MTD and MTDg models in case of high-order dependencies. In the analysis, special attention is paid to transitions from bank- to insurance-product groups, or reverse, as these might indicate paths to convert possible insurance- or bank-only customers into financial-services customers. This topic is of major concern to the IFSP as most of its customers solely possess bank products or insurance products, inhering in the pre-merger history of the group.

3.1. Customer selection

For this study, we had the data warehouse of the IFSP for Belgium at our disposal. The database contains purchase history and profile variables for 3.5 million Belgian customers. To ensure the data is reliable, we only selected customers who were already customer on January 1st, 1999, or who became customer from that date on. This results in about 3 million customers. In order to be able to compare models up to third order, we only selected customers with at least four distinct purchase events. The first three observations of the purchase sequences will not be used to estimate the transition probabilities. The latter ensures that the number of observations is the same for all models (e.g., first-order, second-order and third-order models), so we can compare these models using LL and BIC criterion. 493,608 customers are kept in this way. Customers with one or more missing values on the product group variable, indicating in which product group the customer did buy, were deleted. This leaves us only 308,238 customers or approximately 10% of the original number of sequences. Out of these customers, we randomly selected a sample of 50,000 customers. The fact that only 10% of the customers satisfied the above mentioned conditions we do not see as a problem since (1) the group is still sufficiently large, (2) from a cost/benefit perspective, we assume cross-sell actions (possibly inspired by the discovered acquisition patterns) towards customers having at least three

[2] The information in the acquisition sequences is at the product group level, not at the product level (cf. infra).

Table 3
Nine product groups offered by the IFSP

	Product group	Description
Banking	1	Savings and investments with low interest rates and no time horizon (e.g., savings accounts)
	2	Savings and investments with fixed low to medium interest rates and time horizon (e.g., forward accounts and bonds)
	3	Long term high-risk investment products (e.g., strategic funds)
	4	Medium-risk investment products without time horizon (e.g., structured funds)
	5	Short- and long-term credits (e.g., mortgages and loans)
	6	Checking accounts
Insurance	7	Fire insurance
	8	Car insurance
	9	Other types of insurance (e.g., health, household, accident and life insurance policies)

previous purchase moments to be more profitable than similar actions towards customers having less purchase moments, (3) although a minimum sequence length of only two is necessary for sequence analysis, we find it more valuable to illustrate the parsimony of the MTD model on longer sequences, allowing for the representation of high-order dependencies, (4) the sampling bias due to selecting only customers having at least four distinct purchase moments seems minor (cf. infra, Section 4.1).

3.2. Sequence-construction procedure

For each customer in the sample we created a purchase sequence as follows: In order to keep the number of states feasible, the sequences contain purchase behaviour at the *product group* level, hence not at the *product* level. More states would provide a more detailed picture of the acquisition process and the cross-buying behaviour. A consecutive predictive model based on these product-group level acquisition patterns could only suggest in what *product group* a customer might purchase next. In order to deduce cross-sell actions at a lower level of aggregation, for instance products, sequences should contain purchases at the product level. However, a large number of states results in estimation problems for underrepresented values. Hence, a trade-off had to be made. Following the internal policy of the IFSP, nine product groups are distinguished based on the characteristics of the services. The first six are banking-product groups, whereas the last three are insurance-product groups (see Table 3).

Each sequence consists of subsequent purchase *events*. For example, if a customer first purchases in category 1, next in category 6, then in category 1 and his fourth and last purchase (given a certain observation window) is again in category 1, the customer's acquisition sequence looks like $1 \rightarrow 6 \rightarrow 1 \rightarrow 1$ and has length 4. However, sometimes a customer buys several products at one purchase moment. This might induce two problems.

Firstly, due to administrative reasons, several product purchases at one moment might appear with *several* open dates in the database. We detected this situation in case of short-term loans and mortgages. If a customer took out a loan and his purchase of checking account, card, debt-balance insurance or car insurance appears in the database with an open date within an interval of two weeks, these seemingly different purchase events are reduced to one purchase event. Similarly, if a customer subscribes to a mortgage and purchases a checking account, card, savings account or debt-balance insurance, which appears with an open date within an interval of two months, these purchases are conceived as one purchase event[3] (see Appendix A for more details).

[3] In Belgium, a customer taking out a mortgage is usually tied to the opening of a current account or an insurance policy (although explicit cross-selling is not legal).

This leads us to the second problem: Purchases within several product groups at one purchase event. In a further attempt to keep the number of states low, we reduced combinations of product groups to only one product group. According to the experience of the IFSP and our own opinions, two groups of rules were used. Firstly, we tried to reduce the combination of product groups to the leading product group, i.e., the purchase in the product group triggering the purchases in the other product groups. For instance, whenever there was a mortgage accompanied with another product group the combination was reduced to product group 5. Secondly, when no leading product group is present, we reduce the combination of product groups to the product group, which is

- most difficult to sell, inferred from the purchase frequency distribution across product groups (e.g., high-risk investments; product group 3, preferred to medium-risk investments; product group 4);
- most profitable (e.g., checking accounts; product group 6, are dominated by all other product groups);
- most specific[4] (e.g., product group 9; other types of insurance, is dominated by all other product groups, except for product groups 1; flexible savings accounts, and 6; checking accounts);
- most contact intensive (cf. savings and investments with fixed low to medium interest rates and time horizon; product group 2, preferred to insurance products; product groups 7–9).

These rules result in the following hierarchy of product groups: $5>3>4>2>8>7>9>1>6$. The latter indicates that, for instance, product group 3, long term high-risk investments, dominates product groups 4, 2, 8, 7, 9, 1 and 6, except from product group 5.

Finally, remark that by reducing simultaneous purchases within several product groups to purchases within *one* product group, our focus is not on instantaneous but rather on inter-purchase event cross-buying behaviour. Our model gives insight in what leading, most profitable, most specific or most contact intensive product group a customer, having at least three previous purchase moments, might be interested in, discarding the other product groups the customer might acquire products in at the same next purchase event. Market basket analysis could be applied to examine simultaneous purchase behaviour (Van den Poel et al., 2004). However, problems might arise from underrepresented purchase combinations.

3.3. Characteristics of product-group sequences

In this section, we describe the sequences in the sample along two dimensions: (1) length and (2) distribution of the product groups.

The minimum length of the sequences is four (cf. selection criteria). Fifty percent of all sequences consist of five purchase events or more. Only one quarter of the sequences is six or more purchase events long. The maximum length represents rather an outlier (i.e., 102) as the 90th percentile is only nine.

Looking at the distribution of the product groups over all sequences, product group 2 occurs most frequently (24.3%). However, this is more a reflection of the fixed time horizon of these products, resulting in the need to renew them after the time horizon expires (Larivière and Van den Poel, 2004). The second largest product group is 6 (19.17%), followed by the car insurance product group (13.71%). Product groups 1, 7 and 5 are approximately equal in size (about 10%). It seems that products in product groups 9 (5.94%), 4 (2.09%) and 3 (1.97%) are less often acquired. A possible explanation for this lies in the acquisition difficulty of the investments in product groups 4 and 3 together with the fact that other types of insurance are mostly only needed once (e.g., a customer subscribes only once to a health insurance policy).

[4] Product group 9 is rather heterogeneous (cf. health, household, accident and life policies), which makes marketing actions rather difficult.

4. Results

4.1. Model fit comparison: Markov, MTD and MTDg

We estimated first-order, second-order and third-order Markov, MTD and MTDg models. Fourth-order models were not run, because the number of observations (50,000), does not allow for the estimation of a fourth-order Markov model, and, hence, we would not be able to compare this Markov model with the same order MTD and MTDg model. In our application the random variable X indicates in which product group a customer buys: $N=\{1–9\}$. With $m=9$ the number of independent parameters to estimate increases rapidly. If the order of the Markov model would be $k=4$, the number of parameters would be $m^k(m-1)=9^4(9-1)=52{,}488$ with only 50,000 observations available. Compare this to a similar fourth-order MTD model with only 75 independent parameters to estimate!

We compare the fit of the estimated Markov, MTD and MTDg models using the BIC criterion. To enable this, we did not take into account the first three observations of each purchase sequence, resulting in 143,949 transitions to predict for the first-, second- and third-order models. The model with the lowest BIC value is opted for.

Table 4 summarises our results. As the independence model is worse than any other model, we have evidence that there is dependence between successive purchases of financial services by the customer. Among the fully parameterised Markov models, the second-order model achieves the lowest BIC value (i.e., 441,290). This is even slightly lower than the BIC values for the second-order MTD and MTDg models. Nevertheless, with 647 independent parameters, the Markov model is far less parsimonious than the MTD (73) and MTDg (140) model. Overall, the best result is realized by the third-order MTD model (BIC=433,010). The third-order MTDg model does only slightly worse. Because of the nice interpretation possibilities of the MTDg model, we will interpret the results of the third-order MTD as well as those of the third-order MTDg model.

The parameters of the third-order MTD model are $\lambda_1=0.4626$, $\lambda_2=0.2917$, $\lambda_3=0.2457$ for the lag parameters and

$$Q=\begin{bmatrix}
\mathbf{0.3540} & 0.1963 & 0.0413 & 0.0338 & 0.1616 & 0.1110 & 0.0412 & 0.0451 & 0.0157\\
0.0469 & \mathbf{0.8365} & 0.0165 & 0.0290 & 0.0088 & 0.0094 & 0.0167 & 0.0250 & 0.0113\\
0.0924 & 0.1694 & \mathbf{0.4959} & 0.1758 & 0.0239 & 0.0158 & 0.0090 & 0.0138 & 0.0040\\
0.0766 & 0.2391 & 0.1259 & \mathbf{0.4410} & 0.0267 & 0.0077 & 0.0260 & 0.0486 & 0.0084\\
0.0914 & 0.0162 & 0.0091 & 0.0028 & \mathbf{0.5937} & 0.0835 & 0.0560 & 0.0646 & 0.0828\\
0.2775 & 0.0547 & 0.0144 & 0.0025 & \mathbf{0.4277} & 0.0417 & 0.0622 & 0.0524 & 0.0669\\
0.0389 & 0.0704 & 0.0081 & 0.0106 & 0.1059 & 0.0211 & 0.1426 & \mathbf{0.3861} & 0.2163\\
0.0201 & 0.0487 & 0.0048 & 0.0100 & 0.0435 & 0.0146 & 0.2104 & \mathbf{0.5195} & 0.1285\\
0.0238 & 0.0746 & 0.0033 & 0.0041 & 0.1327 & 0.0287 & 0.2444 & \mathbf{0.3121} & 0.1764
\end{bmatrix} \quad (12)$$

for the transition matrix. Notice that the third-order reduced Markov transition matrix R would be 729×9, whereas the third-order MTD transition matrix Q is only 9×9, making interpretation more straightforward. Please keep in mind that the transition probabilities in (12) only apply to customers having at least three previous purchase events.

The lag parameters indicate a positive but diminishing influence of the purchase history on the customer's purchase at moment t. Knowing in which product group a customer bought last time is already a good estimator of what product group he will buy in next. The $t-3$ purchase still contributes 24.57% to the current state. The positive effects of the lags can also be derived from the transition matrix, as the larger probability values are mostly on the first diagonal (cf. figures in bold indicate largest transition probability from a certain product group).

Table 4
Comparing up to third-order Markov, MTD and MTDg models

Model	Order	LL	BIC	Number of parameters[a]
Independence		−278,030	556,150	8
Markov	1	−230,590	461,330	72
	2	−216,800	441,290	647
	3	−207,250	467,760	4485
MTD	2	−220,670	442,200	73
	3	−216,070	433,010	74
MTDg	2	−220,410	442,490	140
	3	−215,620	433,660	203

[a] Because structural zeros are not taken into account, the number of independent parameters can be lower than when using the formulae of Table 2.

A first look at the transition matrix Q (12) teaches us that if a customer having at least three previous purchase events, bought in a certain product group at time $t-1$ (or $t-2$ or $t-3$), this increases his probability of buying in that product group at moment t (cf. positive effect of the lags). Especially for product group 2, i.e., savings and investments with fixed low to medium interest rates and time horizon, there is a strong positive relationship between the three lags and the probability of buying in product group 2 again at moment t (cf. 0.8365).

Before analysing the transition matrix of the third-order MTD model in detail, we come back to the possible sampling bias due to selecting customers having at least four distinct purchase events (cf. supra). We tested for this sampling bias by creating also a random sample of 50,000 customers having at least two distinct purchase moments—two is the minimum sequence length needed for sequence analysis—(referred to as sample 2) and a random sample of 50,000 customers having at least three distinct purchase events (referred to as sample 3). We compared the best model on sample 2 (e.g. first-order Markov) with the best model on our sample with minimum length 4 (referred to as sample 4) (e.g. third-order MTD model). The main trends are corroborated, but some new relationships appear. Although there are some differences between the transition matrices stemming from samples 4 and 2, these differences mainly originate from the different order of the models rather than from sample bias induced by the requirement in the length of the sequences (2 as opposed to 4). This is reflected by the fact that the transition matrices of a first-order Markov model on sample 4 and of the same model estimated on sample 2 are quite similar. We also compared the best model for sample 4 to the best MTD model on sample 3. Although the second-order Markov model has the lowest BIC value on sample 3, we use the second-order MTD transition matrix as this matrix is easier to compare with the third-order MTD model and the performance difference is rather small (BIC=406,340 for second-order Markov vs BIC=407,240 for second-order MTD). Once more, there are differences in the transition matrices, but these are smaller than was the case for sample 2. There are almost no differences between the transition matrices stemming from applying the second-order MTD model to sample 2 or 4. In conclusion, the sampling bias from selecting only customers having at least four distinct purchase events seems to be minor.

In what follows, we analyse the transition matrix of the third-order MTD model as well as the third-order MTDg transition matrices into more detail. Firstly, we discuss the transition behaviour for the bank product groups, i.e., product groups 1–6, see solid dashed box in (12). Next, we analyse the transition probabilities for the insurance product groups, i.e., product groups 7–9, see dots and dashed box in (12). Special attention is paid to transitions that are more difficult and to transitions from bank product groups to insurance product groups and reverse. Again, please keep in mind that the transition probabilities only apply to customers having at least three previous purchase events. Therefore, in what follows, whenever we speak of 'the customer', we refer to a customer having a purchase history of at least three long.

4.2. Discussion of the transition probabilities for the bank product groups

If a customer bought at moment $t-3$, $t-2$ or $t-1$ in *product group 1*, i.e., savings and investments with fixed interest rates and no time horizon, he has a medium probability to buy again in that product group (35.40%) at moment t. However, that customer still has a probability of 64.60% to buy in another product group, hence inducing cross-buying behaviour. He has a probability of 19.63% to buy a savings or investment product with fixed low to medium interest rates and time horizon (cf. product group 2), which can be explained by the same low risk nature of the product group. Besides, the customer might also purchase a short- or long-term credit (i.e., product group 5) or a checking account (i.e., product group 6). The lag-specific transition matrices of the MTDg3 model (see Appendix B) are not very different from the MTD transition matrix. This means that the transition probabilities are relatively independent from the timing of the purchase in product group 1. In general, we clearly observe a relationship between product groups 1, 2, 5 and 6. Apparently, the purchase of a simple savings product (i.e., product group 1) mostly precedes later purchases of savings and investment products at low to medium risk but with time horizon (i.e., product group 2) (cf. 19.63% vs 4.69%) as well as the subscription of short-term or long-term loans (i.e., product group 5) (cf. 16.16% vs 9.14%). On the contrary, the customer is more likely to open a checking account (i.e., product group 6) before taking a savings or investment product at low interest rate without time horizon (i.e., product group 1) (cf. 27.75% vs 11.10%). This indicates that product group 6 is lower in the financial-services need hierarchy than product group 1 and that the latter is positioned below product groups 2 and 5.

With a repurchase probability of more than 80%, *product group 2* (i.e., savings and investment products with fixed low to medium interest rates and time horizon) is not really an initiator of cross-product group purchases. Only a small part of the customers who bought in product group 2 over the last three purchase events cross-buy in product group 1. However, there is a larger flow from product group 1 to 2 than reverse (cf. 19.63% vs 4.69%).

What if the customer bought a product in *product group 3*, i.e., long-term high-risk investments, during the last three purchase events? Then he is very likely to buy in one of the investment-related categories; product groups 3, 4 and 2 (mentioned along descending transition probabilities). The repeat product-group purchase probability is highest when the customer acquired a product in product group 3 at his previous purchase event (cf. 55.15% vs 38.63% vs 44.55%). However, there is no linear diminishing positive effect of the lags on the repeat-purchase probability. Looking for cross-buying behaviour, we observe high transition probabilities to product group 4 (17.58%) and product group 2 (16.94%).

The repeat product-group repurchase probability for *product group 4* (i.e., medium-risk investments without time horizon) is 44.10%. This implies that the majority is distributed over other product groups. Just like product group 3, there is a strong relationship with the other investment product groups. However, whereas from product group 3 it is as likely to buy products from product group 2 or 4 (cf. 16.94% and 17.58%), from product group 4 it is more likely to buy product group 2 than 3 (cf. 23.91% vs 12.59%).

If a customer acquired a product from *product group 5*, i.e., short- and long-term credits, he has 59.37% to repurchase in that product group. Whether this customer will acquire products from other product groups later severely depends on when he took out the loan. There is almost no cross-buy possibility when the credit was subscribed three purchases ago, as the repeat purchase probability exceeds 80% (see lag-3 transition probability from product group 5 to 5, see Appendix B). However, when the credit subscription was one of his last two purchase events, he might acquire products from product groups 1, 6, 9, 8 and 7. These cross-product group transition probabilities are also lag sensitive. Once again, the additional interpretational possibilities of the MTDg model come into play. In general, loans are closely related to savings and investments with low interest rates and no time horizon, and to checking accounts as banking product groups, but also to all insurance product groups, leaving possibilities to transform bank-only customers into financial-services customers (see Section 4.4).[5]

What if a customer opened a checking account (i.e., *product group 6*) over his last three purchase events? This customer has a very high cross-buying tendency as the repeat-purchase probability is only 4.17%! The explanation for this is twofold: Firstly, once a customer has a checking account, his need for an account to put money on that is available at any time is satisfied. Only occasionally, the opening of another checking account is useful. Secondly, checking accounts are rather basic services after which other products are acquired (Stafford et al., 1982; Kamakura et al., 1991). Hence, product group 6 is located at the bottom of the pyramid of financial needs or in other words, the opening of a checking account mostly comes first in the acquisition pattern. The bigger stream from product group 6 to product groups 1 (cf. 27.75% vs 11.10%), 2 (cf. 5.47% vs 0.94%), 5 (cf. 42.77% vs 8.35%), 7 (cf. 6.22% vs 2.11%), 8 (cf. 5.24% vs 1.46%) and 9 (cf. 6.69% vs 2.87%) than reverse reflects this. Therefore, although checking accounts are not profitable in themselves, they might be valuable and even necessary in inducing purchases in more profitable groups. Product group 5 (42.77%) and 1 (27.75%) are within the bounds of possibility. The relationship between product groups 1, 5 and 6 was already apparent from investigating the transition matrices for product groups 1 and 5. Other smaller cross-buy possibilities (i.e., ±6%) are observed for product groups 9, 7, 2 and 8. Just like product group 5, product group 6 triggers the subscription to insurance policies (cf. infra, Section 4.3). Unlike product group 5, checking accounts encourage purchases of savings and investments at low to medium interest rates and with time horizon (i.e., product group 2). Apparently, some of the customers opening a checking account (27.75%) invest their money not needed on short-term into product group 1, still having it available anytime they want to, whereas a much smaller part (5.47%) puts it into product group 2, accepting the unavailability of the invested amount over the fixed time horizon. Analogous to product group 5, the cross-product group transition probabilities are lag-specific. The probability to purchase in product group 5 increases the longer the elapsed time since the opening of the checking account (cf. 47.73% vs 68.71% vs 81.20%). Product groups 5 and 1 are always the best cross-buying possibilities, but the ordering of the other cross-buying product groups is dependent on the lag.

Fig. 1 gives an overview of the most common flow between bank-product groups. The arrows indicate which product group is usually acquired before acquiring the other product group, i.e., is situated earlier in the acquisition pattern. For instance, $6 \rightarrow 1$, indicates that there is a bigger stream from 6 to 1 than reverse. Only transitions of minimum 5% are represented. The percentages indicate the highest transition probability to the product group. The loop in product group 2 depicts its very high repeat-purchase probability. In general, we distinguish two product group areas, which represent more or less two types of acquisition behaviour. One customer can show both types of behaviour over his relationship with the IFSP. Some customers might exhibit only one type. For instance, customers in area 1 (i.e. product groups 1, 2, 5 and 6) might start their relationship with a rather basic product, i.e., a checking account (i.e., product group 6). Subsequently, they acquire a simple savings account (i.e., product group 1), some form of credit (i.e., product group 5) or a savings and investment product with low to medium risk and time horizon (i.e., product group 2). If their second purchase event involved the opening of a simple savings account, their next purchase might be in product groups 2 or 5. A second type of behaviour is rather investment related (cf. area 2; product groups 1–4). These customers make investments in one of the investment product groups (i.e., 2, 3 or 4) after which they might invest in the same product group (e.g. product group 2) or one of the remaining investment groups. They might even open a simple savings account later on. As such, in area 2

[5] It is possible that the customer is already a financial-services customer, because he subscribed to an insurance policy before t.

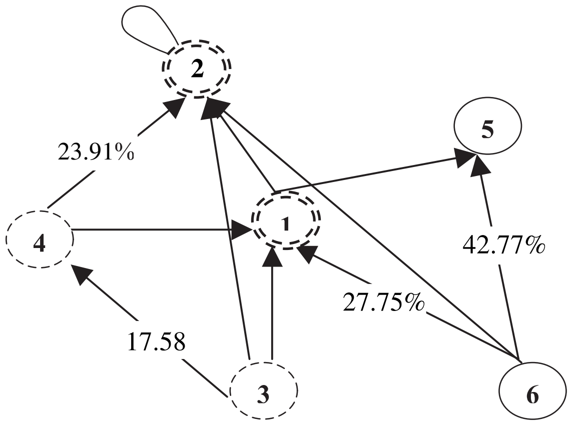

Fig. 1. Flow chart for bank product groups.

there is mostly a stream towards product group 1, whereas in area 1, there are several arrows departing from product group 1.

4.3. Discussion of the transition probabilities for the insurance product groups

A customer who took out an insurance policy over his last three purchase occasions has a high probability to buy in one of the remaining insurance product groups. Product groups 7 (i.e., fire insurance) and 9 (i.e. other types of insurance) are both characterized by a low repeat purchase probability, which is inherent in the nature of these insurance products. Customers do not very often feel the need to purchase fire insurance, as most of them do not move very frequently nor buy several homes. Similarly, health, household, accident and life insurance policies are typical products, which are subscribed only once. Conversely, car insurance policies are subscribed several times, which results in a repeat-purchase probability of over 50%. Fig. 2 illustrates the most common flow between the insurance product groups. The arrows indicate which product group is usually acquired before the other product group. The double-sided arrow indicates the lag-dependency of the flows between product groups 7 and 9. Once again, this illustrates the added value of the MTDg model.

4.4. How to convert customers into real financial-services customers?

As the IFSP only recently became a financial-services company, one of its major concerns is cross-selling bank-product groups to insurance-only customers and insurance-product groups to bank-only customers.

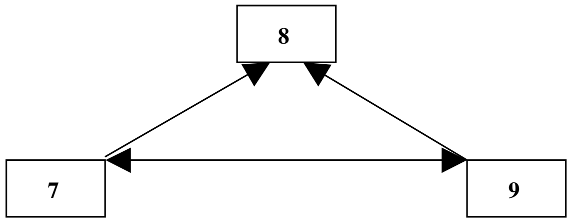

Fig. 2. Flow chart for insurance product groups.

Only a small part of its customer list owns products from bank-product groups as well as from insurance-product groups. Therefore, an analysis of the sparse cross-buying patterns from bank to insurance product groups or reverse might shed light on possible strategies to increase the number of real financial-services customers. Still, only a predictive study could give certainty on whether these patterns are indicative for future bank-insurance purchase behaviour and hence, could guide marketing cross-sell strategies.

What bank to insurance cross-buying behaviour do we observe for customers having at least three previous purchase events? Most bank-insurance patterns depart from bank product groups 5, 6, 4 and 1 (mentioned along descending order of probabilities). In what insurance product group the possibly bank-only customer subscribes a policy, depends on when he acquired an account in one of these bank product groups. In general, we observe a positive effect of the recency of the purchase within product groups 5, 6, 4 or 1 on the insurance purchase probability. For instance, a customer is more likely to acquire an insurance product when he took out a credit (i.e., product group 5) over his last two purchase events than when his loan subscription dates back three purchase events ago (cf. low transition probabilities to insurance product groups for lag 3). If the credit subscription was the customer's last purchase, he might be inclined to subscribe car insurance policy (i.e., product group 8) or other types of insurance (i.e., product group 9) (cf. 10.20% and 10.09%, respectively), whereas a purchase within other types of insurance is more likely for a customer who obtained credit two purchase events ago. Remark that the biggest flow from bank to insurance product groups occurs for customers who take out a short- or long-term credit and subsequently subscribe car insurance or other types of insurance at the next purchase event (cf. 10.20% and 10.09% are the highest transition probabilities). Although checking accounts are not profitable in themselves, they might present an opportunity to cross-sell and even to stimulate bank-only customers to also purchase their insurance policies from the IFSP. If a bank-only customer opened a checking account at his last purchase event, he is very willingly to buy a credit or simple savings account. However, some of these customers also have a small possibility to subscribe other types of insurance (cf. 8.03%) or car insurance policy (cf. 7.18%). If the customer opened a checking account two purchase events ago, he has a higher probability to cross-buy fire insurance (cf. 8.85%) or other types of insurance (cf. 7.46%) than to subscribe car insurance (cf. 3.13%). Other smaller conversion probabilities lie in cross-selling car insurance to customers whose last purchase event involved purchases within product groups 1 or 4. Fig. 3 illustrates possible paths to convert a bank-only customer into a financial-services customer. The line style indicates when (i.e., for which lags) the cross-buying patterns are present. Only transitions of minimum 4% are taken into account. The percentages refer to the highest transition probability considering the lag-specific transition probabilities to the product group.

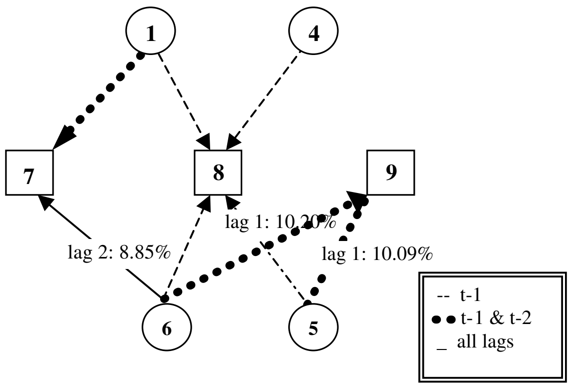

Fig. 3. Possible paths to convert bank-only customers into financial-services customers.

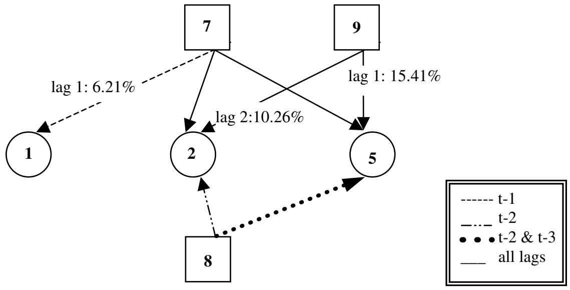

Fig. 4. Possible paths to convert insurance-only customers into financial-services customers.

Which insurance to bank cross-buying behaviour do possibly insurance-only customers having at least three previous purchase events, display? In general, these customers are always likely to cross-buy in product groups 5 and 2, no matter what insurance policy they recently subscribed to. However, this cross-buying intention is bigger for a customer who subscribed to fire insurance or other types of insurance (i.e., product groups 7 and 9), than for a customer who recently took out car insurance (i.e., product group 8). Once more, the lag-specific transition matrices provide additional insight and might be used to decide which bank-product group to cross-sell. For instance, if a customer subscribed to another type of insurance at his last purchase event, the manager might recommend taking out a loan (i.e., product group 5). However, if the other types of insurance was purchased two purchase events ago, the manager could try to cross-sell a savings and investment product with fixed low to medium interest rates and time horizon (i.e., product group 2) besides product group 5. Nevertheless, as the transition probabilities from product group 9, lag 2, to product groups 5 and 2 are almost identical, it might be preferable to promote product group 5 rather than product group 2, as the latter triggers almost no cross-product group purchases. Fig. 4 illustrates possible paths to persuade insurance-only customers to also buy their bank products at the IFSP.

In conclusion, note that the sum of the differences between (a) the lag-specific transition probabilities from a given insurance product group to a given banking product group and (b) the lag-specific transition probabilities from the same banking product group to the same insurance product group (e.g., $7 \rightarrow 1$ is 6.21% for lag 1, $1 \rightarrow 7$ is 4.43% for lag 1 → 6.21–4.43%) is positive. Hence, it might be easier to transform an insurance customer into a financial-services customer than a bank customer.

5. Summary, limitations and directions for further research

In this paper, we proposed the MTD and MTDg models as alternatives for the commonly used Markov model to parsimoniously analyse sequences. We applied up to third-order Markov, MTD and MTDg models on financial-services acquisition sequences. Comparing the BIC values of the models, the third-order MTD model had the best fit, immediately followed by the MTDg model. Moreover, with only 74 parameters to estimate, the third-order MTD model is far more parsimonious than the corresponding fully parameterised Markov model (4,458 parameters). Although Raftery (1985a) introduced the MTD model to parsimoniously estimate *high-order* Markov chains, the parsimony of the MTD model is also relevant in case of a *large number of possible values for the random variable X* (i.e., states). The latter will often be the case in marketing contexts. The transition matrix of the third-order MTD model was mainly used for investigation of cross-buying patterns, which might help to outline cross-sell strategies. The limited size of the MTD transition matrix ($m \times m$) instead of ($m^k \times m$) for the Markov transition matrix, facilitated the interpretation of the results, enormously. From this small size transition matrix it is easy to create flow charts illustrating how customers move between product groups. This insight might be of great value to product managers. The charts are also valuable to identify possible paths to cross-sell or even to convert bank-only of insurance-only customers into financial-services customers. Furthermore, for product groups 5, 7, 8 and 9, we found the lag-specific transition matrices of the MTDg model to better represent the transition behaviour than the MTD transition matrix. Hence, the ability to represent situations where the relationship between each lag and the state at moment t differs, is a quite valuable feature of the MTDg model.

This study is not free of some limitations: Firstly, the results of our study only apply to customers having at least three previous purchase events. For customers, having less purchase events, other models should be built and compared. Secondly, we did not satisfy the discrete-time assumption of the discrete Markov, MTD and MTDg model. The acquisition of financial services was not measured at constant discrete moments in time (for instance every two months), because the latter would not give a total view on the sequence of purchases. For example, if a customer opens a checking account in the first month, then subscribes to a loan and a fire insurance in the second month, a discrete time sequence would be like $6 \rightarrow 5$ and 7. The latter should be reduced to one category to keep the number of states limited to nine. In our approach, the sequence contains all purchase events, no matter when they did occur: $6 \rightarrow 5 \rightarrow 7$. However, according to one of the reviewers, this limitation is minor as not satisfying the discrete-time assumption is usual practice in some fields such as behaviour analysis and psychology. Thirdly, as Kamakura et al. (2003) indicated, the use of single-source data, e.g. customer transaction database information on the ownership of products at the IFSP, is not optimal for identifying cross-sell possibilities as it overlooks the possibility that the customer already owns the product at a competitor. Fourthly, another limitation concerns the nature of the sequences investigated. To what extent are the cross-buying patterns discovered really customer driven and not the result of the marketing strategy of the financial-services provider (Baesens et al., 2004)? Given distinct levels of intensity of marketing campaigns in the past and certain levels of flows towards the product(group), conclusions can be drawn on whether it concerns natural or induced acquisition behaviour and whether current transition probabilities represent cross-sell opportunities (see Table 5). On the one hand, a small flow to a product group (from other product groups) indicates that there is no natural tendency to cross-buy this product group. However, it might be possible to induce this behaviour by more (if few marketing campaigns occurred in the past) and/or more efficient marketing campaigns. On the other hand, current big flows to a product group (from other product groups) might indicate natural cross-buying behaviour, if few marketing actions promoted this cross-buying in the past or if, when the marketing-action intensity for this product group decreases, the flow to the product group stays equal (purely natural behaviour) or only slightly decreases (natural and induced behaviour). In this study, we are not able to make a distinction between natural and induced cross-buying behaviour, neither

Table 5
Natural vs induced cross-buying behaviour

Observed flow	Past marketing actions	
	A few	A lot
Small	*No natural acquisition behaviour present.* Maybe possible to induce behaviour with more intensive and/or more efficient marketing actions than in the past	*No natural acquisition behaviour.* Induced behaviour might be possible through more efficient marketing campaigns than in the past
Big	*Real natural acquisition behaviour.* Maybe possible to stimulate this natural behaviour by more marketing campaigns with as a result an even bigger flow to the product (group). If the flow stays the same, marketing actions are redundant	*Natural acquisition behaviour?* If less marketing actions, but at least as efficient as in the past, and the flow to the product (group): (1) Stays as big $\rightarrow$ natural behaviour (2) Decreases $\rightarrow$ partially natural, partially induced behaviour (3) Disappears $\rightarrow$ induced behaviour

can we evaluate whether current cross-sell possibilities epitomize opportunities, as the IFSP does not possess detailed historical marketing-action information. Moreover, the multichannel nature of the marketing communication towards the financial-services customer (e.g. personal contact during branch visit, direct-mail, tele-marketing, e-marketing, etc.) largely hampers the evaluation of the efficiency of (past) marketing actions.

Fifthly, using product groups rather than single products results in discovering rather general cross-buying patterns and hence identification of possible cross-sell paths at the product group level. However, in order to keep the number of states limited, some reduced picture of the total product assortment has to be accepted. One could opt for a selection of products or, like we did, to work with product groups. Moreover, we believe that the use of modelling techniques for cross-sell purpose will never make the knowledge/experience of the sales force redundant. In this case, we assume the salesperson has enough customer knowledge to recommend the customer the right product out of the product group proposed by the model.

Several paths are open for further research: Firstly, the unreduced transition matrix of the third-order MTD model reveals that some transitions are not fully third order. As the variable length Markov chain model (VLMC) (Weinberger et al., 1995; Bühlmann and Wyner, 1999), one of the alternative models for high-order Markov chains, supposes that only a part of the structure of the data is of full *k*th order, the integration of the VLMC model with the MTD model seems valuable. After all, Berchtold and Raftery (2002) have shown that the MTD model and the VLMC model are complementary. Unfortunately, to date, the VLMC model is not suited to multiple-sequence analysis. Secondly, we overlooked the influence of customer heterogeneity on the cross-sell possibilities. Heterogeneity could be accounted for in two ways: One could estimate the third-order model on several subsamples based on a covariate like the age of becoming customer. Alternatively, covariates could be incorporated into the MTD model through the addition of a supplementary term. Finally, replicating the study at other companies, possibly in other countries/cultures could enhance the external validity of the results.

Acknowledgements

The authors would like to thank the anonymous financial-services company for providing the data. Next, we extend our thanks to Bart Larivière, Ph.D. candidate at Ghent University, for carrying out much of the preparatory data processing on the data warehouse of this company. Moreover, we would like to thank André Berchtold for making available the software via his website to estimate the Markov, MTD, and MTDg models <http://www.andreberchtold.com>. Finally, we would like to thank (1) Ghent University for funding the PhD project of Anita Prinzie (BOF Grant no. B00141), and (2) the Flemish Research Fund (FWO) for providing the funding for the computing equipment to complete this project (Grant no. G0055.01.N).

Appendix A

See Tables 6 and 7.

Table 6
Detecting purchases with distinct open dates in the database, which are in reality only one purchase event: short-term credit case

Short-term credit in combination with	Product groups bought in at one purchase event
Checking account	5+6
Checking account+cards (e.g., giro checks and credit cards)	5+6+6 → 5+6
Debt-balance insurance	5+9
Car insurance	5+8

Table 7
Detecting purchases with distinct open dates in the database, which are in reality only one purchase event: mortgage case

Mortgage in combination with	Product groups bought in at one purchase event
Checking account	5+6
Checking account+cards (e.g., giro checks and credit cards)	5+6+6 → 5+6
Savings account	5 + 1
Debt-balance insurance	5+9
Fire insurance	5+7
Fire insurance+checking account	5+7+6
Fire insurance+checking account+cards (e.g., giro checks and credit cards)	5+7+6+6 → 5+6+7
Fire insurance+savings account	5+7+1
Fire insurance+debt-balance insurance	5+7+9

Appendix B. Lag-specific transition matrices for third-order MTDg model

Lag 1

$$Q = \begin{bmatrix} \mathbf{0.3695} & 0.1784 & 0.0420 & 0.0345 & 0.1412 & 0.1031 & 0.0443 & 0.0658 & 0.0212 \\ 0.0579 & \mathbf{0.8142} & 0.0166 & 0.0258 & 0.0170 & 0.0077 & 0.0131 & 0.0379 & 0.0098 \\ 0.0774 & 0.1295 & \mathbf{0.5515} & 0.1706 & 0.0222 & 0.0125 & 0.0062 & 0.0262 & 0.0038 \\ 0.0825 & 0.2265 & 0.1230 & \mathbf{0.4178} & 0.0332 & 0.0043 & 0.0252 & 0.0822 & 0.0053 \\ 0.1184 & 0.0224 & 0.0143 & 0.0049 & \mathbf{0.4773} & 0.0845 & 0.0752 & 0.1020 & 0.1009 \\ 0.2907 & 0.0494 & 0.0218 & 0.0016 & \mathbf{0.3769} & 0.0547 & 0.0527 & 0.0718 & 0.0803 \\ 0.0621 & 0.0521 & 0.0117 & 0.0089 & 0.0869 & 0.0196 & 0.1582 & \mathbf{0.3762} & 0.2242 \\ 0.0387 & 0.0382 & 0.0092 & 0.0083 & 0.0360 & 0.0141 & 0.1913 & \mathbf{0.5602} & 0.1038 \\ 0.0439 & 0.0624 & 0.0063 & 0.0039 & 0.1541 & 0.0327 & 0.2396 & \mathbf{0.2977} & 0.1595 \end{bmatrix}$$

Lag 2

$$Q = \begin{bmatrix} \mathbf{0.3464} & 0.2012 & 0.0422 & 0.0271 & 0.1616 & 0.1163 & 0.0534 & 0.0273 & 0.0244 \\ 0.0215 & \mathbf{0.8582} & 0.0168 & 0.0336 & 0 & 0.0196 & 0.0281 & 0.0065 & 0.0156 \\ 0.1233 & 0.2451 & \mathbf{0.3863} & 0.1858 & 0.0023 & 0.0285 & 0.0125 & 0.0105 & 0.0057 \\ 0.0732 & 0.2458 & 0.1171 & \mathbf{0.4605} & 0.0123 & 0.0232 & 0.0372 & 0.0105 & 0.0148 \\ 0.0552 & 0.0154 & 0.0043 & 0 & \mathbf{0.6871} & 0.0843 & 0.0518 & 0.0324 & 0.0695 \\ 0.2448 & 0.0617 & 0.0039 & 0.0031 & \mathbf{0.4500} & 0.0422 & 0.0885 & 0.0313 & 0.0746 \\ 0.0303 & 0.0946 & 0.0082 & 0.0145 & 0.1294 & 0.0421 & 0.1620 & \mathbf{0.3499} & 0.1690 \\ 0.0021 & 0.0756 & 0 & 0.0096 & 0.0418 & 0.0240 & 0.2315 & \mathbf{0.4658} & 0.1495 \\ 0.0000 & 0.1026 & 0.0012 & 0.0020 & 0.0928 & 0.0306 & 0.2583 & \mathbf{0.3230} & 0.1896 \end{bmatrix}$$

Lag 3

$$Q = \begin{bmatrix} \mathbf{0.3317} & 0.2272 & 0.0413 & 0.0399 & 0.1815 & 0.1255 & 0.0232 & 0.0297 & 0 \\ 0.0590 & \mathbf{0.8451} & 0.0170 & 0.0303 & 0 & 0 & 0.0143 & 0.0244 & 0.0100 \\ 0.1125 & 0.1866 & \mathbf{0.4455} & 0.1594 & 0.0659 & 0.0105 & 0.0150 & 0 & 0.0045 \\ 0.0689 & 0.2608 & 0.1591 & \mathbf{0.4653} & 0.0248 & 0 & 0.0119 & 0 & 0.0092 \\ 0.0668 & 0 & 0.0027 & 0 & \mathbf{0.8120} & 0.0749 & 0 & 0.0064 & 0.0371 \\ 0.2997 & 0.0510 & 0.0102 & 0.0025 & \mathbf{0.5047} & 0.0130 & 0.0505 & 0.0336 & 0.0349 \\ 0 & 0.0940 & 0 & 0.0107 & 0.1272 & 0.0039 & 0.0778 & \mathbf{0.4485} & 0.2378 \\ 0.0049 & 0.0359 & 0.0008 & 0.0137 & 0.0698 & 0.0049 & 0.2291 & \mathbf{0.4767} & 0.1643 \\ 0.0146 & 0.0613 & 0 & 0.0076 & 0.1174 & 0.0168 & 0.2390 & \mathbf{0.3447} & 0.1986 \end{bmatrix}$$

References

Abbott, A., 1995. Sequence analysis: New methods for old ideas. Annual Review of Sociology 21, 93–113.

Abbott, A., Hrycak, A., 1990. Measuring resemblance in sequence data: An optimal matching analysis of musicians' careers. American Journal of Sociology 96 (1), 144–185.

Adke, S.R., Deshmukh, S.R., 1988. Limit distribution of a high-order Markov chain. Journal of the Royal Statistical Society B 50, 105–108.

Agrawal, R., Srikant, R., 1995. Mining sequential patterns. In: 11th Int. Conf. Data Engineering, ICDE.

Baesens, B., Verstraeten, G., Van den Poel, D., Egmont-Petersen, M., Van Kenhove, P., Vanthienen, J., 2004. Bayesian network classifiers for identifying the slope of the customer-lifecycle of long-life customers. European Journal of Operational Research 156 (2), 508–523.

Berchtold, A., 2001. Estimation in the mixture transition distribution model. Journal of Time Series Analysis 22 (4), 379–397.

Berchtold, A., Raftery, A.E., 2002. The mixture transition distribution model for high-order Markov chains and non-Gaussian time series. Statistical Science 17 (3), 328–356.

Berger, A.N., 2003. The efficiency effects of a single market for financial services in Europe. European Journal of Operational Research 150 (3), 466–481.

Berger, A.N., Demsetz, R.S., Strahan, P.E., 1999. The consolidation of the financial services industry: Causes, consequences, and implications for the future. Journal of Banking & Finance 23 (2–4), 135–194.

Bollerslev, T., Chou, R.Y., Kroner, K.F., 1992. ARCH modeling in finance: A review of the theory and empirical evidence. Journal of Econometrics 52, 5–59.

Buckinx, W., Van den Poel, D., forthcoming. Customer base analysis: Partial defection of behaviorally-loyal clients in a non-contractual FMCG retail setting. European Journal of Operational Research.

Bühlmann, P., Wyner, A.J., 1999. Variable length Markov chains. Annals of Statistics 27 (2), 480–513.

Chatfield, C., Lemon, R.E., 1970. Analysing sequences of behavioural events. Journal of Theoretical Biology 29, 427–445.

Clarke, Y., Soutar, G.N., 1982. Consumer acquisition patterns for durable goods: Australian evidence. Journal of Consumer Research 8, 456–459.

Cohen, A., Ivry, R.I., Keele, S.W., 1990. Attention and structure in sequence learning. Journal of Experimental Psychology: Learning, Memory & Cognition 16 (1), 17–30.

Feick, L., 1987. Latent class models for the analysis of behavioral hierarchies. Journal of Marketing Research 24 (2), 174–186.

Felvey, J., 1982. Cross-selling by computer. Bank Marketing, 25–27.

Hauser, J.R., Urban, G.L., 1986. The value priority hypotheses for consumer budget plans. Journal of Consumer Research 12 (4), 446–462.

Hebden, J.J., Pickering, J.F., 1974. Patterns of acquisition of consumer durables. Oxford Bulletin of Economics and Statistics 36 (2), 67–94.

Hopp, W., 1987. A sequential model of R&D investment over an unbounded time horizon. Management Science 33 (4), 500–508.

Jonz, J., 1989. Textual sequence and 2nd-language comprehension. Language Learning 39 (2), 207–249.

Kamakura, W.A., Ramaswami, S.N., Srivastava, R.K., 1991. Applying latent trait analysis in the evaluation of prospects for cross-selling of financial services. International Journal of Research in Marketing 8, 329–349.

Kamakura, W.A., Wedel, M., de Rossa, F., Mazzon, J.A., 2003. Cross-selling through database marketing: A mixed data factor analyzer for data augmentation and prediction. International Journal of Research in Marketing 20 (1), 45–65.

Kasulis, J.J., Lush, R.F., Stafford, E.F., 1979. Consumer acquisition patterns for durable goods. Journal of Consumer Research 6, 47–57.

Katz, R.W., 1981. On some criteria for estimating the order of a Markov chain. Technometrics 23 (3), 243–249.

Katz, L., Proctor, C.H., 1959. The concept of configuration of interpersonal relations in a group as a time-dependent stochastic process. Psychometrika 24, 317–327.

Larivière, B., Van den Poel, D., 2004. Investigating the role of product features in preventing customer churn, by using survival analysis and choice modeling: The case of financial services. Expert Systems with Applications 27 (2), 277–285.

Le, N.D., Martin, R.D., Raftery, A.E., 1996. Modeling flat stretches, bursts, and outliers in time series using mixture transition distribution models. Journal of the American Statistical Association 91, 1504–1515.

Lipman, D., Pearson, W., 1984. Rapid and sensitive protein similarity searches. Science 22, 1435–1441.

Logan, J.A., 1981. A structural model of the high-order Markov process incorporating reversion effects. Journal of Mathematical Sociology 8, 75–89.

Lush, R.F., Stafford, E.F., Kasulis, J.J., 1978. Durable accumulation: An examination of priority patterns. In: Hunt, K.H. (Ed.), Advances in Consumer Research, vol. 5. Association for Consumer Research, Ann Arbor.

MacDonald, I.L., Zucchini, W., 1997. Hidden Markov and Other Models for Discrete-Valued Time Series. Chapman and Hall, London.

Martin, R.D., Raftery, A.E., 1987. Non-Gaussian state-space modelling of nonstationary time-series—robustness, computation, and non-Euclidean models—comment. Journal of the American Statistical Association 82 (400), 1044–1050.

Mayo, M.C., Qualls, W.J., 1987. Household durables goods acquisition patterns: A longitudinal study. In: Anderson, P., Wallendorf, M. (Eds.), Advances in Consumer Research, vol. 14. Association for Consumer Research, Ann Arbor, MI.

McBrearty, S., 1988. The Sagoan–Lupemban and middle stone-age sequence at the Muguruk site. World Archaeology 19 (3), 388–420.

McFall, J., 1969. Priority patterns and consumer behavior. Journal of Marketing 33, 50–55.

Paroush, J., 1965. The order of acquisition of consumer durables. Econometrica 33 (1), 225–235.

Pegram, G.G.S., 1980. An autoregressive model for multilag Markov chains. Journal of Applied Probability 17, 350–362.

Pyatt, F.G., 1964. Priority Patterns and the Demand for Household Durable Goods. Cambridge University Press, Cambridge.

Raftery, A.E., 1985a. A model for high-order Markov chains. Journal of the Royal Statistical Society, Series B 47 (3), 528–539.

Raftery, A.E., 1985b. A new model for discrete-valued time series: Autocorrelations and extensions. Rassegna di Metodi Statistici ed Applicazioni (3–4), 149–162.

Raftery, A.E., 1993. Change point and change curve modeling in stochastic processes and spatial statistics. Journal of Applied Statistical Science (1), 403–423.

Raftery, A.E., Banfield, J.D., 1991. Stopping the Gibbs sampler, the use of morphology and other issues in spatial statistics. Annals of the Institute of Statistical Mathematics 43 (1), 32–43.

Raftery, A.E., Tavaré, S., 1994. Estimation and modelling repeated patterns in high order Markov chains with the mixture transition distribution model. Applied Statistics 43 (1), 179–199.

Sabherwal, R., Robey, D., 1995. Reconciling variance and process strategies for studying information system development. Information Systems Research 6 (4), 303–327.

Schwarz, G., 1978. Estimating the dimension of a model. Annals of Statistics (6), 461–464.

Srikant, R., Agrawal, R., 1996. Mining sequential patterns: Generalizations and performance improvements. IBM Almaden Research Center Paper.

Srivastava, R., Shocker, A.D., 1987. Strategic challenges in the financial services industry. In: Pettigrew, A. (Ed.), The Management of Strategic Change. Basil Blackwell, Oxford.

Stafford, E.F., Kasulis, J.J., Lush, R.F., 1982. Consumer-behavior in accumulating household financial assets. Journal of Business Research 10 (4), 397–417.

Van den Poel, D., Larivière, B., 2004. Customer attrition analysis for financial services using proportional hazard models. European Journal of Operational Research 157 (1), 196–217.

Van den Poel, D., De Schamphelaere, J., Wets, G., 2004. Direct and indirect effects of retail promotions on sales and profits in the do-it-yourself market. Expert Systems with Applications 27 (1), 53–62.

Weinberger, M.J., Rissanen, J.J., Feder, M., 1995. A universal finite memory source. IEEE Transactions on Information Theory 41 (3), 643–652.

Wong, C.S., Li, W.K., 2000. On a mixture autoregression model. Journal of the Royal Statistical Society B 62, 95–115.

Xu, J.G., Agrawal, R., 1996. Membrane separation process analysis and design strategies based on thermodynamic efficiency of permeation. Chemical Engineering Science 51 (3), 365–385.