Modelling the purchase dynamics of insurance customers using Markov chains

Jean-Francois Bozzetto, Leilei. Tang, Lyn C. Thomas*, Stephen Thomas,

School of Management,

University of Southampton

*** corresponding author ; Prof Lyn C Thomas**

School of Management, University of Southampton, Southampton, SO17 1 BJ, UK.

Tel 44-(0)2380597718

Fax 44-(0)2380593844

Email: l.thomas@soton.ac.uk

Modelling the purchase dynamics of insurance customers using Markov chains

Abstract

This paper considers how various types of Markov chains can be used to help forecast the purchase behaviour of customers. The models are used in a case study of the purchase behaviour of the customers of a major insurance company.As well as looking at the impact of relaxing the first order Markov and time homogeneity assumptions which are usually used in Markov chain models, the paper also looks at models based on mover-stayer ideas and ones which enlarge the state space by including the type of purchase as well as the time of purchase. One important aspect of long term customer relationships such as those which occur in the insurance and assurance industry is the impact of changes in the economy. The final section show how these can be incorporated into Markov chain models and how they can make a significant difference to the quality of the predictions.

Keywords: Consumer Behavior; Data mining; Probability Models

1 Introduction

The advent of data warehouses with their ability to store and analyse large amounts of data, has made it possible for large financial organisations to become better at forecasting what their customers are likely to do and hence much more discriminating in their relationships with their customers and the marketing of products to them. This enables the organisation to have much more sophisticated customer relationship management strategies, especially in the service industries (Gronroos 2000) and to develop more accurately the customer lifetime value (CLV). CLVs have proved to be a useful concept in customer relationship management (CRM) as Dwyer (1989) showed in his work on customer retention and customer migration. It is particularly useful in the financial service area of banking and insurance where the length of the relationship between customer and service provider can almost be the customer's actual lifetime. One crucial element in building such lifetime value models is to estimate the probability of a customer making a purchase in each future time period. This paper develops a range of such purchase models.

Markov chain models have proved useful in modelling the dynamics of a random process in many different contexts; see the examples in the classic texts of Feller (1957) and Iosifescu (1980) and the financial application in Kijima (2003). In the context of customer behaviour, Cyert (1962) was the first to develop a Markov chain model of customer's repayment behaviour and although there have only been a limited number of subsequent applications it was the advent of the data warehouses that revived interest in such modelling in the consumer credit risk area (Ho et al 2004, Trench et al 2003). Schneiderjans and Lock (1994) used Markov chain models to model the marketing aspects of customer relationship management in the banking environment. Pfeifer and Carraway (2000) provided a very clear tutorial on how to use Markov chain modelling to calculate customer lifetime values and give an example where the states of the system describe the recency of a purchase. However this was a hypothetical example and used the simplest type of the Markov chain process. It did though point out that once one has built a successful Markov chain model one can extend the model to a Markov decision process which will optimize the decisions to be made as part of customer relationship management.

This paper looks at a case study of using Markov chain models to estimate the purchase dynamics of the customers of an international insurance organisation. As indicated previously, getting a valid model of when and how often a customer makes a purchase is a crucial element in estimating their lifetime values and hence in building CLV models and supporting CRM decisions. The models considered concentrate on the number of purchases so is strongly related to the Frequency part of the RFM (Recency, Frequency, Monetary Value) framework (Shepherd 1995). However the paper also seeks to show the variety of Markov chain models available apart from the basic first order stationary chains which have appeared in the literature so far. It also indicates the ways of testing whether the models give good fits and examines their predictive accuracy. Perhaps the most important class of models developed are those where the transition probabilities depend on the current state of the economy. Not only does this provide a way of modelling the well established theory that consumer purchase patterns are affected by changes in the economy it also allows the forecasts of future customer purchases, and so of CLV, to be consistent with an organisation's forecasts of the likely changes in the economy. This is particularly important in sectors like finance and insurance where customer relationships last for a long time during which the economy is likely to go through several cycles.

Section two recalls the definitions of the different types of Markov chains and how one can use goodness of fit tests to check the appropriateness of a particular model type to explain the data. Section three describes the data on customers and their purchase patterns held in the data warehouse of the international insurance company, to be used in this case study. Section four looks at the results of building the simplest possible Markov chain model – time homogeneous and first order Markov – and uses these to determine what is an appropriate base transition time period. The next section investigates extensions of this basic model by allowing for models with time inhomogeneity and less restrictive Markovity assumptions. Section six investigates alternative Markov models with slightly more complex state spaces, while section seven deals with the models where the Markov chain transition probabilities are assumed to be functions of the underlying economy. In these models the transition probabilities are derived using logistic regression approaches. Finally we draw some conclusions on the pros and cons of the different models.

2 The Markov framework

Markov chains have proved ubiquitous stochastic processes because their simplicity belies their power to model a variety of situations. Formally, we define a discrete time $\{t_0, t_1, \ldots, t_u, \ldots; u \in U\}$ and a discrete state space $S = \{1, 2, \ldots, E\}$ Markov chain as a stochastic process $\{X(t_u)\}_{u \in U}$ with the property that $u \in U$ and $i, j \in E$

$$\begin{aligned}\mathbf{P}\left[X(t_{u+1}) = j \mid X(t_0) = s_0, X(t_1) = s_1, \ldots, X(t_u) = i\right] &= \mathbf{P}\left[X(t_{u+1}) = j \mid X(t_u) = i\right] \\ &= q_{ij}(t_u, t_{u+1})\end{aligned}$$

where $q_{ij}(.,.)$ is called the probability transition matrix. This is the Markov property and such stochastic processes are called first order Markov chains.

If $(\pi(t_u))' = (\pi_1(t_u),...,\pi_E(t_u))$ describes the probability distribution of the states of the process at time t_u, the Markov property implies that the distribution at time t_{u+1} can be obtained from that at time t_u by

$$\pi_j(t_{u+1}) = \sum_{i \in S} \pi_i(t_u) q_{ij}(t_u, t_{u+1})$$

This extends to a m-stage transition matrix so that

$$\pi_i(t_{u+m}) = \sum_{j \in S} \pi_i(t_u) q_{ij}{}^m(t_u, t_{u+m})$$

where

$$q_{ij}^m(t_u, t_{u+m}) = \sum_{s_1, s_2, .. s_{m-1}} q_{is_1}(t_u, t_{u+1}) \prod_{r=1}^{r=m-2} q_{s_r s_{r+1}}(t_{u+r}, t_{u+r+1}) q_{s_{m-1}j}(t_{u+m-1}, t_{u+m}) \text{, for } m > 2$$

If the time periods between the t_u are constant, then the Markov chain is time homogeneous or stationary provided

$$q_{ij}(t_u, t_{u+1}) = q_{ij} \qquad \forall\, t_u, t_{u+1}, i, j$$

Given a set of data, for $T+1$ time periods u = 0,1,2….T, Anderson and Goodman (1957) describe how to obtain the maximum likelihood estimators of the transition probabilities of a Markov Chain model of the data. Let $n_u(s_0, s_1, ..., s_k)$ be the number of data points which exhibit the sequence

$$X(t_u) = s_0,\ X(t_{u+1}) = s_1, ...,\ X(t_{u+k}) = s_k$$

Define

$$n(s_0, ..., s_k) = \sum_{0 \le u \le T-k} n_u(s_0, ..., s_k) \quad \text{and} \quad N(s_0, ..., s_k) = \sum_{0 \le u \le T-k-1} n_u(s_0, ..., s_k)$$

as the number of times the sequence of states $(s_0, s_1, ..., s_k)$ occurs at any time in the data sample of histories, (where N ignores the last period) then the maximum likelihood estimates of the transition probabilities for the Markov chain are

$$\hat{q}_{ij}(t_u, t_{u+1}) = \frac{n_u(i,j)}{n_u(i)}, \qquad i, j \in S, \qquad 0 \le u \le T-1$$

If one assumed that the Markov chain was stationary, then the estimate become

$$\hat{q}_{ij} = \frac{n(i,j)}{N(i)}, \qquad i, j \in S$$

One can weaken the Markov property and require the information about the future is not all in the current state, but is in the current and the last state of the process. This is called a second order Markov chain and formally it satisfies the condition

$$\begin{aligned}\mathbf{P}\left[X(t_{u+1}) = k \mid X(t_0) = s_0, ..., X(t_{u-2}) = s_{u-2}, X(t_{u-1}) = i, X(t_u) = j\right] &= \mathbf{P}\left[X(t_{u+1}) = k \mid X(t_{u-1}) = i, X(t_u) = j\right] \\ &= q_{jk}(t_{u-1}, t_u, t_{u+1})\end{aligned}$$

This is equivalent to the process being a first order Markov chain but with state space $S \times S$. The concept can be generalized to defining k^{th}-order Markov chains for any k, though of course, the state space and the size of the transition probability matrices goes up exponentially as k increases. The maximum likelihood estimators in the second order case are

$$\hat{q}_{ij,jk}(t_u, t_{u+1}) = \hat{q}_{ijk}(t_{u-1}, t_u, t_{u+1}) = \frac{n_{u-1}(i,j,k)}{n_{u-1}(i,j)}$$

with comparable definitions for higher orders. For a second order stationary Markov chain, the estimators become

$$\hat{q}(ij, jk) = \hat{q}(i,j,k) = \frac{n(i,j,k)}{\sum_{0 \le u \le T-2} n_u(i,j)} = \frac{n(i,j,k)}{N(i,j)}$$

To check whether a stationary Markov chain describes the data adequately, Anderson and Goodman (1957) made the analogy with contingency tables and so used Pearson goodness of fit chi-squared statistics to test the various hypotheses. Firstly one can check the stationarity of the process. For time t_u this corresponds to the hypothesis that the transition probabilities at that time are the same as those if the process was stationary. This corresponds to hypothesis $H^T(u)$ with chi-square statistic $X^2(u)$ where

$$H^T(u): \quad q_{ij}(t_u, t_{u+1}) = q_{ij} \qquad \forall i, j \in S$$

And

$$X^2(u) = \sum_{j=1}^{E} \sum_{t=1}^{E} \frac{\left[n_{ij}(t_u, t_{u+1}) - n_i(t_u) \frac{n(i,j)}{N(i)} \right]^2}{n_i(t_u) \frac{n(i,j)}{N(i)}}$$

The transition matrix of each Markov chain has E rows and E columns and so appears to have $E(E-1)$ independent entries. However it may contain structural zeros i.e. state movements which are not possible or not allowed and if there are $r(i)$ of these in row i, then the degrees of freedom of the chi-square test is $\mathrm{E(E\text{-}1)} - \sum r(i)$.

This is essentially a diagnosis of where there may be some non-stationarity in the process but the true test of stationarity is that these hypothesis hold at all times t_u, which we label H^T with corresponding statistic X^2 where

$$H^T: \quad q_{ij}(t_u, t_{u+1}) = q_{ij} \qquad \forall i, j \in S \quad \forall u = 0, \ldots, T-1$$

and

$$X^2 = \sum_{u=0}^{T-1} X^2(u)$$

which has $(T-1)\left[\mathrm{E(E\text{-}1)} - \sum r(i)\right]$ degrees of freedom.

To check the Markovity, so that the process can be described by a stationary Markov chain one looks at the hypothesis H^M and corresponding chi-square statistic Y^2 where

$$H^M : q(1,i,j) = q(2,i,j) = \ldots = q(E,i,j), \quad i,j \in S$$

$$Y_i^2 = \sum_{k=1}^{M} \sum_{j=1}^{M} \left[\frac{\left[n(k,i,j) - N(k,i)\frac{n(i,j)}{N(i)} \right]^2}{N(k,i)\frac{n(i,j)}{N(i)}} \right]$$

Y^2 has $E(E\text{-}1)^2$ degrees of freedom if there are no structural zeros but could be far less otherwise.

One can generalize the stationary Markov model by allowing it to be second, third or k^{th} order stationary Markov rather than first order Markov. The hypothesis H^M is essentially a test of first order against second order. One way of testing whether a second order model is more appropriate than a third order model is to recall that a second order Markov chain is essentially a first order on a state space $S \times S$. Similarly, to check if a r$^{\text{th}}$ order chain is suitable, we can check if it is first order on the r-fold product space Sx…..xS. We can apply the standard Markov test in this case and end up by checking the hypothesis

$$H_0^r : q(1,i_1,\ldots,i_r,j) = q(2,i_1,\ldots,i_r,j) = \ldots = q(E,i_1,\ldots,i_r,j), \text{ for all } i_1,\ldots,i_r \in S$$

If there is non-stationarity it could be caused by several reasons. It could be seasonal, in that the time periods are representing different times of the year, and the transition matrices in a given season may be assumed to be all the same, but different to the transition matrices in other seasons. The most general form of non-stationarity would occur if the transition matrix is assumed to be different in each time period. Although this model gives maximum flexibility, it is not very useful as a forecasting tool unless we can relate the differences to other variables for which one has forecasts.

One final extension is the idea that the population whose dynamics the process is seeking to describe is not homogenous and in fact groups in the population do not move at all. Thus the parameters for a time homogenous version of this 'mover-stayer' model are $s_i, \quad i=1,\ldots,E$ which represent the proportion of the population who "stay" in state i (i.e. do not ever change state) and p_{ij} which are the probabilities of a 'mover' going from state i to state j. Movers make up $(1-\sum_i s_i)$ of the population and Frydman (1985) gives the Maximum Likelihood estimators for these parameters.

3 Data on insurance product purchases

In the subsequent sections, these Markov chain ideas are used to build models which help forecast the future purchase behaviour of customers with an insurance company. Although Markov chain models of customer behaviour have been around for many years in the credit risk context (Cyert and Thompson 1962) and such models have recently been made more complex (Trench, Hand and Hill, Ho and al 2004), the use of Markov chain models in customer relationships is much more limited (Pfeifer and Carraway 2000) .

Our models are built using the information on customers and their purchase behaviour, which is held in the database of the direct marketing channel of an international insurance company. The data covered a period from July 1975 until July 2003, but was only a complete record of the transactions in the period January 1999 to July 2003. There were just under 50,000 customers in the database during this four and half year period, and all their purchases of products, their financial payments as part of the product purchase, their (and the firm's) profit from the product and their termination of products were recorded in the database. There were 20 variables which described the status of the customer and 43 variables detailing the product and its performance. The products could be divided into four main groupings: and the appropriate details kept on each type. This paper concentrates on the purchasing of new products. By definition, anyone in the database must have made at least one purchase and during this period there were 5890 purchases of a second product, 1175 purchases of a third product, 137 purchases of a fourth product and 35 purchases of a fifth or further product. Given the small numbers involved with three or more further purchases, we concentrate on forecasting times for the second (one further) purchase and third or higher (2 or more further) purchases are likely to occur.

One aspect our modelling investigates is whether there is a relationship between purchasing and the prevailing economic conditions. Traditional consumer demand analysis focuses on the relation between the prices of goods and consumers' incomes, while saving models include variables such as interest rates, wealth, personal income and consumer sentiment. Here we chose variables to reflect the attractiveness of financial investments and the general economic and investment climate. The UK economic variables considered are Consumer Prices, Consumer Confidence Index, Unemployment Rate, FTSE All Share Index, and Bank of England Base Interest Rate. Transformations of these variables are considered in order to conform with the macro economic literature, to avoid the problems of non-stationary time series and to have variables that relate to the way consumers perceive the economic conditions. In the light of this we chose the following variants of five economic measures of the economy's impact on the consumer

- Stock Market returns: quarterly difference in log of the FTSE100. This variable gives the yield from stock market and also reflects the fact that a buoyant stock market may encourage purchasing of financial products.

- Consumer prices: annual difference of the consumer price index. This represents the price inflation felt by customers and high levels may deter customers buying savings products.

- Confidence index: the index level (difference between those who are more and those who are less confident about the future of the economy) is used because this is a stationary process. It remained in negative territory throughout this period.

- Unemployment rate: yearly difference in unemployment rate (unemployed as percentage of the population available for work). This reflects the changes in jobs available for consumers and may add to the information about the business cycle that is in the confidence index.

- Interest rate: The bank of England LIBOR rate is used. It usually impacts on the customers through the mortgage repayment rate and hence disposable income. It also reflects the opportunity cost of switching savings from bank deposits to financial products.

It is important to test the validity of the models built in an unbiased way. In order to do this we split the sample of 50,000 cases into a development sample consisting of 70% of the cases and a hold out sample which is the remaining 30% of the sample. The models are built and the parameters estimated on the development sample and their ability to forecast is checked on the completely independent holdout sample. This is the easiest way to produce unbiased forecasts and is possible because of the large sample available.

4 Basic Markov model

The simplest model we consider - our basic model – assumes that the purchase process will follow a time-homogeneous first order Markov chain. The states for such a simple model are the number of purchases made by the consumer. Since there is only data on the firm's customers, and not on the general population, everyone in the data set has made at least one purchase. Thus the lowest state would be 1 purchase, and the others could be 2,3,4 purchases etc. However, as was alluded to in the last section, since only 2% of the purchases are fourth, fifth or higher level purchases, we will combine states 3,4,5 etc into the state 3+ in order to develop a more robust model. Thus we concentrate on a three state model with states 1, 2 and 3+ purchases. This implies that the purchase behaviour for the second purchase is allowed to be different from that for the third or higher purchases.

The second decision to be made is what time periods to use. Although the data on a purchase gives a precise purchase date, the infrequency of purchases and the fact that the database is only formally updated and archived each month means we will only consider one month or multiples of one month as the time interval of a transition period. We investigate which of these is best by working out the Pearson goodness of fit statistics for the different time intervals. If the time interval is one month, the maximum likelihood estimate of the transition matrix is

$$\boldsymbol{Q}_1 = \begin{vmatrix} 0.997 & 0.003 & 0.000 \\ - & 0.994 & 0.006 \\ - & - & 1.000 \end{vmatrix}$$

So the transitions over a three month period are given by multiplying three copies of it to give

$$\boldsymbol{Q}_1^{(3)} = \begin{vmatrix} 0.991 & 0.009 & 0.000 \\ - & 0.982 & 0.018 \\ - & - & 1.000 \end{vmatrix}$$

where $\mathrm{Q}_{\mathrm{m}}^{\mathrm{n}}$ is the n^{th} fold transition matrix when the basic period is m months and $\left(q_m^n\right)_{ij}$ is the $(i,j)^{th}$ entry

Note that although this is a 3x3 matrix there are in fact only 3 probabilities to be estimated q_{11}, q_{12} and q_{22} since three of the entries are structural zeros, and three are defined by the stochastic matrix condition that the rows must add up to 1. For a model with a three month basic transition period, the maximum likelihood estimators lead to a transition matrix

$$\boldsymbol{Q}_3^{(1)} = \begin{vmatrix} 0.992 & 0.008 & 0.000 \\ - & 0.983 & 0.017 \\ - & - & 1.000 \end{vmatrix}$$

which is slightly different to $\boldsymbol{Q}_1^3$ even though they measure transitions over the same time period.

In order to decide on the best base time period we compare the time homogeneity of a stationary model for each possible base time period. What we are testing is the hypothesis H^T that the purchase process can be explained by a time homogeneous first order Markov chain as compared with a time nonhomogeneous first order Markov chain with the same base time period. Table 1 gives the X^2 values for the different possible transition periods m for $m = 1,2,3,4,6,12$.

Transition period	X^2	dof	$p-value$ lower tail	χ^2 95%	χ^2 99.99%
1	477.96	141	1	169.71	212.14
2	371.47	69	1	89.39	121.44
3	221.85	45	1	61.65	89.07
4	242.44	33	1	47.40	72.02
6	155.95	21	1	32.67	53.95
9	83.98	12	1	21.03	39.13
12	78.83	9	1	16.62	33.72

Table 1. Stationary tests for different base periods

In no case is the hypothesis H^T close to being accepted but we did not expect this to happen. We use the X^2 values (compared with the corresponding χ^2) to get some feel of how much time homogeneity there is with each base time period. The results suggest that the 3 monthly time interval is as competitive as the others, as it is the only one where the X^2 value is lower than that for larger base time periods. Since one is used to quarters as an appropriate time period for economic measurements this is the time period we favour. Transition time periods of one or two months lead to less robust models, while longer time periods makes short term forecasting impossible.

To check how good this 3-month base period model is, we look at how closely it satisfies the time homogeneity and the Markovity tests outlined in Section 2. Table 2 gives these results. Neither the Markovity not the time homogeneity hypothesis can be accepted though for some time periods, the time homogeneous transition matrix is not too distant an approximation. There is no pattern of when time homogeneity is a good or poor approximation to the actual transition matrix.

Hypothesis	X^2	dof	$p-value$ lower tail	χ^2 95%	χ^2 99.99%
$H^T(1)$	4.565	3	0.79	7.81	21.10
$H^T(2)$	6.760	3	0.92	7.81	21.10
$H^T(3)$	29.197	3	1.00	7.81	21.10
$H^T(4)$	5.694	3	0.87	7.81	21.10
$H^T(5)$	7.954	3	0.95	7.81	21.10
$H^T(6)$	9.330	3	0.97	7.81	21.10
$H^T(7)$	19.981	3	1.00	7.81	21.10
$H^T(8)$	12.695	3	0.99	7.81	21.10
$H^T(9)$	29.574	3	1.00	7.81	21.10
$H^T(10)$	25.059	3	1.00	7.81	21.10
$H^T(11)$	0.846	3	0.16	7.81	21.10
$H^T(12)$	43.741	3	1.00	7.81	21.10
$H^T(13)$	7.366	3	0.94	7.81	21.10
$H^T(14)$	3.380	3	0.66	7.81	21.10
$H^T(15)$	6.280	3	0.90	7.81	21.10
$H^T(16)$	9.432	3	0.98	7.81	21.10
H^T	221.855	45	1.00	61.66	89.07
H^M	15.966	1	1.00	3.84	15.13

Table 2. Stationary and Markovity tests – 1st order 3-month based Markov chain

Another way of assessing a model is to look at its predictive ability. To do this we take the holdout sample and compare the estimates of the cumulative number of purchases over different time horizons in the future with the actual numbers of purchases made. Let $n_m(t,1)$ be the number of customers who enter the holdout sample by making their first purchase in the t^{th} period from the start, and only make one purchase in that first period, where the base period is m months.. Let $n_m(t,2)$ and $n_m(t,3)$ be the same definitions but for those who make two or three purchases in the first month they enter the holdout sample. Period 0 represent those already in the data base at its start. The expected number of purchases in the first k periods

(km months) predicted by the model with base period m months can be obtained by first calculating the number of customers $c_m(t,i)$ who have made i cumulative purchases by time t. These satisfy

$$c_m(t+1,1) = c_m(t,1)q^1_{m\,11} + n_m(t,1)$$
$$c_m(t+1,2) = c_m(t,1)q^1_{m\,12} + c_m(t,2)q^1_{m\,22} + n_m(t,2)$$
$$c_m(t+1,3) = c_m(t,1)q^1_{m\,13} + c_m(t,2)q^1_{m\,23} + c_m(t,3) + n_m(t,2)$$

The predicted number of non-initial purchases by time t is then given by

$$\mathrm{C_m}(t) = 0\,c_m(t,1) + 1\,c_m(t,2) + 2\,c_m(t,3)$$

The results for the 3-month model are given in Figure 1 where Figure 1a gives the actual and forecasted number of cumulative purchases and Figure 1b gives the cumulative forecasting error - the difference between the actual and the forecasted number of cumulative purchases. This will be used as a bench mark against which to compare future models.

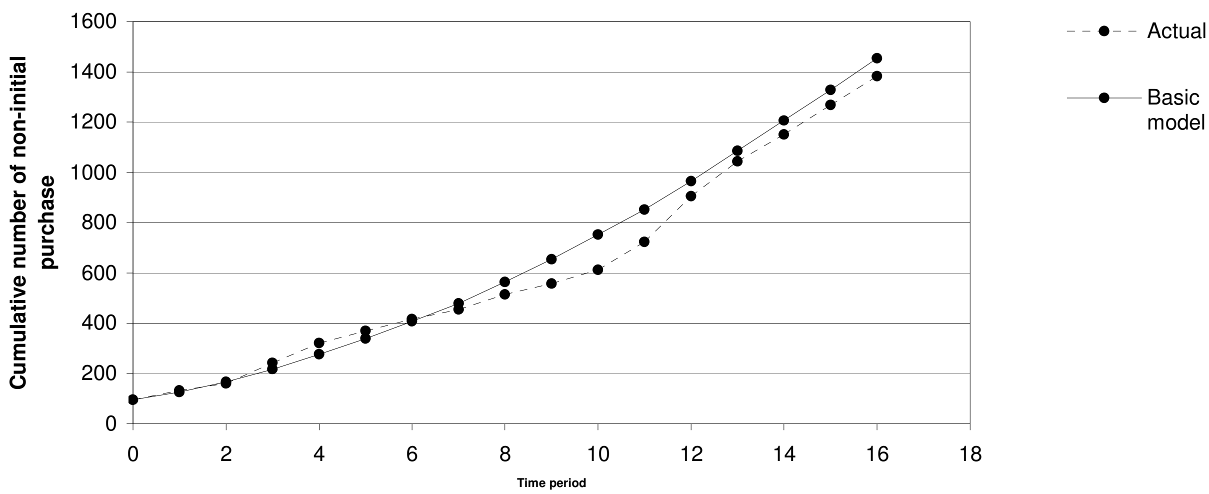

Figure 1.a Actual and basic model (3-month period) forecast for cumulative number of non-initial purchase

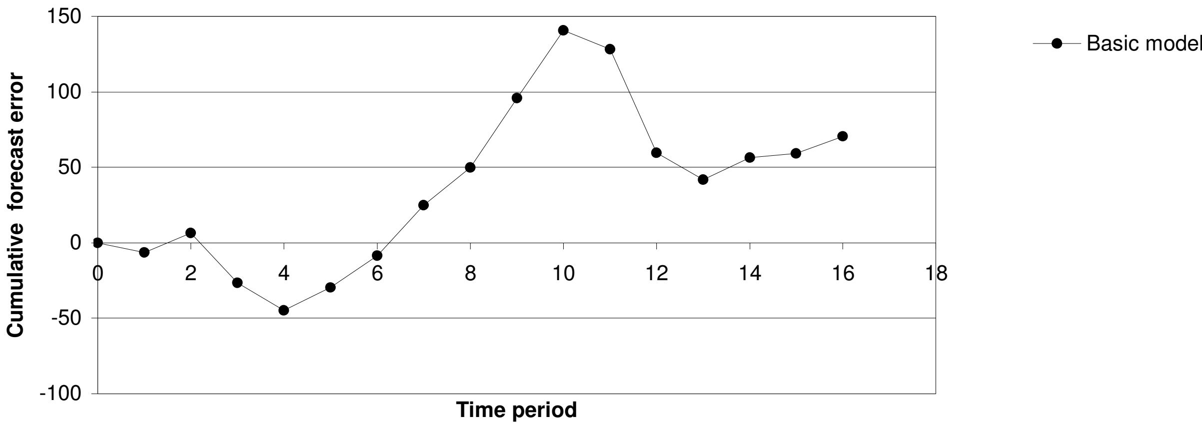

Figure 1.b Cumulative forecast error using basic model

5 Higher order and non-stationary Markov versions of the Basic model

As was suggested in section two, one can generalize the Markov model in two ways, either by weakening the Markovity assumption or weakening the time homogeneity assumption. In this section we look at both these extensions. Having determined that quarters (3-monthly intervals) are the most satisfactory base transition time period we will continue to use this as our time period in the generalizations considered hereafter and so drop the subscript m from our notation. So hereafter the transition probabilities are denoted q_{ij} .

Firstly we consider whether the first order Markov assumption on the model is satisfactory or should we go for a second order Markov chain. In a second order Markov chain the state space is given by $(i.j)$ where i is the number of purchases at the start of the last quarter and j is the number of purchases at the start of this quarter. The possible states are then (1,1), (1,2), (1,3+), (2,2), (2,3+), (3+,3+) and the model constructs a 6x6 Markov transition matrix on these states. One can redefine the states if one wishes as the number of purchases to date and the number of purchases in the last quarter, since the state (2,2) would mean two purchases to date but none in the last quarter. The transition matrix for this model is given by

$$Q_{3\,(2^{nd}\,order)} = \begin{vmatrix} 0.992 & 0.008 & - & - & - & - \\ - & - & - & 0.979 & 0.021 & - \\ - & - & - & - & - & 1.000 \\ - & - & - & 0.984 & 0.016 & - \\ - & - & - & - & - & 1.000 \\ - & - & - & - & - & 1.000 \end{vmatrix}$$

To check if this is a better fit than the first order chain, one applies the chi-square test on the H^M hypothesis. The results were given in Table 2 and suggest that the first order hypothesis is not really valid in this case. One might then ask whether the second order Markov property is appropriate and to do this one compares the Markov chain on the six states with the Markov chain where the states are (i, j, k) describing the situation where the cumulative number of purchases is currently k , was j at the start of the last period and i at the start of the period before. In this case with the forced monotonicity in the way the process moves between states there are only ten possible states (1,1,1), (1,1,2), (1,1,3+), (1,2,2), (1,2,3+), (1,3+,3+), (2,2,2), (2,2,3+), (2,3+,3+) and (3+,3+,3+). The corresponding transition matrix is given by

$$Q_{3\,(3^{rd}\,order)} = \begin{vmatrix} 0.992 & 0.008 & 0.000 & - & - & - & - & - & - & - \\ - & - & - & 0.980 & 0.020 & - & - & - & - & - \\ - & - & - & - & - & 1.000 & - & - & - & - \\ - & - & - & - & - & - & 0.992 & 0.008 & - & - \\ - & - & - & - & - & - & - & - & 1.000 & - \\ - & - & - & - & - & - & - & - & - & 1.000 \\ - & - & - & - & - & - & 0.983 & 0.017 & - & - \\ - & - & - & - & - & - & - & - & 1.000 & - \\ - & - & - & - & - & - & - & - & - & 1.000 \\ - & - & - & - & - & - & - & - & - & 1.000 \end{vmatrix}$$

Table 3 gives the hypothesis tests for the time homogeneity and Markovity of the second order model. The results indicate again that the data does not really support such hypotheses and so one might repeat the process for the third order model. We will not do so here but sufficient to say that the 15 state fourth order model is both difficult to understand and has non-robust estimators of the transition probabilities.

Hypothesis	X^2	dof	$p-value$ lower tail	χ^2 95%	χ^2 99.99%
$H^T(1)$	5.637	4	0.77	9.49	23.51
$H^T(2)$	43.240	4	1.00	9.49	23.51
$H^T(3)$	7.663	4	0.90	9.49	23.51
$H^T(4)$	9.036	4	0.94	9.49	23.51
$H^T(5)$	7.642	4	0.89	9.49	23.51
$H^T(6)$	21.018	4	1.00	9.49	23.51
$H^T(7)$	11.489	4	0.98	9.49	23.51
$H^T(8)$	28.187	4	1.00	9.49	23.51
$H^T(9)$	22.077	4	1.00	9.49	23.51
$H^T(10)$	2.615	4	0.38	9.49	23.51
$H^T(11)$	39.270	4	1.00	9.49	23.51
$H^T(12)$	8.609	4	0.93	9.49	23.51
$H^T(13)$	2.809	4	0.41	9.49	23.51
$H^T(14)$	11.157	4	0.98	9.49	23.51
$H^T(15)$	11.553	4	0.98	9.49	23.51
H^T	232.001	56	1.00	74.47	104.13
H^M	17.439	1	1.00	3.84	15.13

Table 3. Stationary and Markovity tests - 2nd order Markov chain

Clearly going to a higher order of Markovity will always give a slightly better fit but the question is whether it is worth the complication. To test this we look again at the forecasts of the cumulative number of purchases of the first, second and third order Markov chain against the actual number of purchases on the holdout sample. Figure 2 suggests that as one would expect, going to a higher order decreases the cumulative error on average (but not necessarily on every time horizon). However the improvements are small and suggest that even if the first order hypothesis is not statistically valid, not much is lost in predictive accuracy by using it.

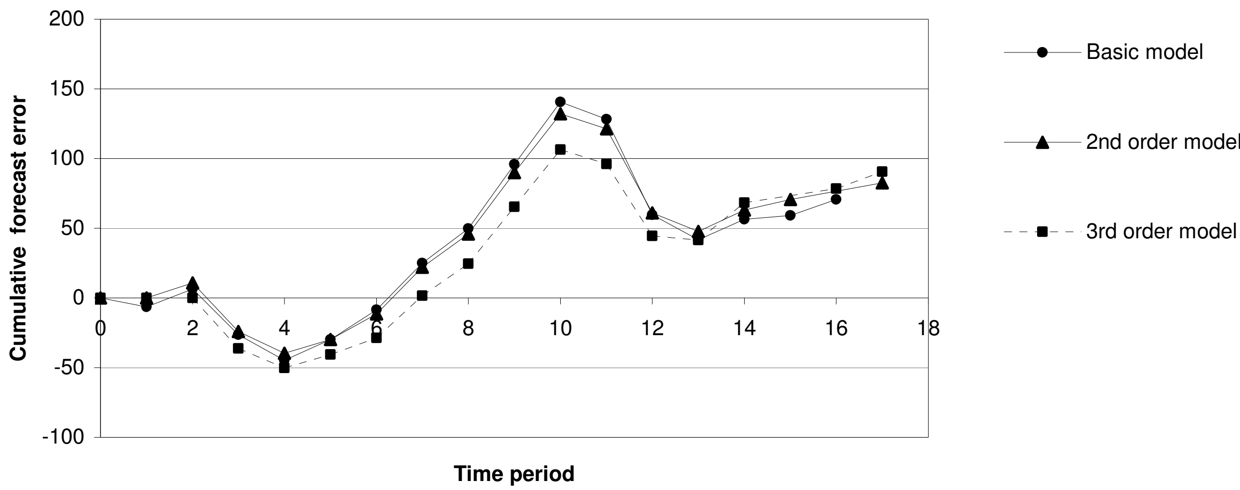

Figure 2. Cumulative forecast error using 2nd and 3rd order models

The second way of generalizing the basic quarterly model is to assume it is non-time homogeneous. Here we consider two levels of extensions in time non-homogeneity. Firstly we assume the process is seasonal, so that there are four transition matrices that describe the dynamics of the process. In the first quarter of the year the transition probabilities are given by $Q(1)$, in the second quarter by $Q(2)$ and so on. At the start of the next year though the transition probabilities in the first quarter are again modelled by $Q(1)$. Using our data we found the best estimates for the $Q(i)$s were

$$Q(1)=\begin{vmatrix} 0.997 & 0.003 & 0.000 \\ - & 0.994 & 0.006 \\ - & - & 1.000 \end{vmatrix}$$

$$Q(2)=\begin{vmatrix} 0.997 & 0.003 & 0.000 \\ - & 0.991 & 0.009 \\ - & - & 1.000 \end{vmatrix}$$

$$Q(3)=\begin{vmatrix} 0.997 & 0.003 & 0.000 \\ - & 0.997 & 0.003 \\ - & - & 1.000 \end{vmatrix}$$

$$Q(4)=\begin{vmatrix} 0.998 & 0.002 & 0.000 \\ - & 0.997 & 0.003 \\ - & - & 1.000 \end{vmatrix}$$

The most notable feature is the increase in third purchases in the second quarter of the year and a corresponding drop in such purchases in the final quarter of the year. The former might be due to existing savers making another tax free investment just before the end of the tax year on April 5. Although we have no evidence for this, it is a possible interpretation of this variation.

The most general form of time non-homogeneity is to allow a different matrix of transition probabilities for each time period, so there will be matrices $Q(1), Q(2),..., Q(T)$. Table 4 gives the values of the transition probabilities for the 16 different quarters in the time horizon available. Clearly this must be a better fit than the seasonal and the stationary models. The figures in Table 1 are essentially the result of testing this against the null hypothesis of complete time homogeneity and it is clear one would need to reject that hypothesis whatever the base time period. Table 2 gave the results for the 3-month base period model and again overall one cannot statistically support time homogeneity. The effects of the non-stationary models on the predictions of the number of purchases in the hold out sample are given by Figure 3 and the results there again confirm their superiority over the stationary models.

Period	End month	q_{11}	q_{12}	q_{13}	q_{22}	q_{23}
1	3	0.989	0.011	0.000	0.945	0.055
2	6	0.993	0.007	0.000	0.992	0.008
3	9	0.987	0.013	0.000	0.978	0.022
4	12	0.990	0.010	0.000	0.986	0.014
5	15	0.992	0.008	0.000	0.988	0.012
6	18	0.993	0.007	0.000	0.012	0.008
7	21	0.995	0.005	0.000	0.988	0.012
8	24	0.993	0.006	0.001	0.985	0.015
9	27	0.995	0.005	0.000	0.995	0.005
10	30	0.996	0.004	0.000	0.995	0.005
11	33	0.989	0.011	0.000	0.983	0.017
12	36	0.987	0.013	0.000	0.970	0.030
13	39	0.988	0.012	0.000	0.979	0.021
14	42	0.994	0.006	0.000	0.989	0.011
15	45	0.991	0.009	0.900	0.980	0.020
16	48	0.993	0.007	0.000	0.973	0.027

Table 4. Transition matrices $Q(1), Q(2), \ldots, Q(16)$

$Q(1)$: Study period: 01/07/1999 to 01/10/1999

$Q(16)$: Study period: 01/04/2003 to 01/07/2003

The results show the non-stationary models can deal with the blip in purchases around quarters 9 to 11, and so the errors for both non-stationary models are superior to the stationary one, over the first three years. The fully non-stationary model has lower errors than the seasonal one over that period but in the longer period it is worse than the basic stationary model as well as the seasonal one.

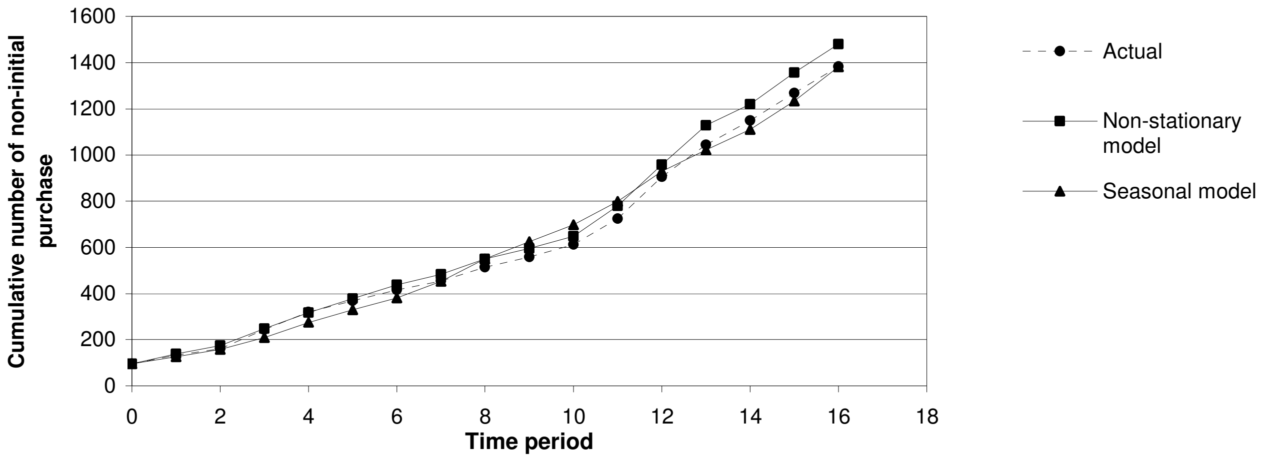

Figure 3.a Actual and non-stationary model (3-month period) forecast for cumulative number of non-initial purchase

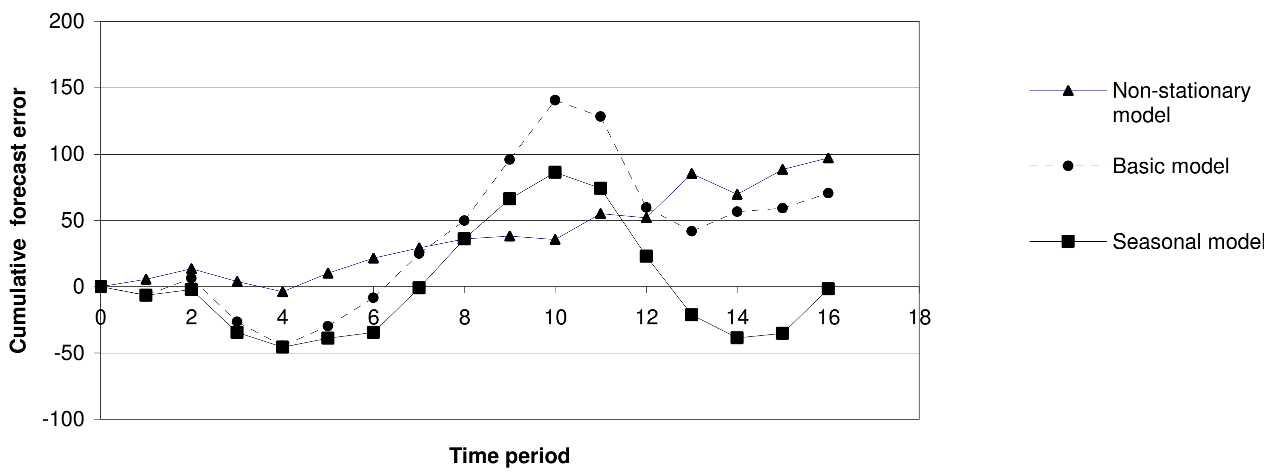

Figure 3.b Cumulative forecast error using a non-stationary model

6 Alternative Markov models

The flexibility of the Markov approach allows a number of variations on these models by segmenting the population and building different Markov processes on each segment or alternatively by adding some extra information to the simple state used in the basic model. One of the most successful applications of segmenting the population is the mover stayer idea which was first used for industrial mobility (Blumen et al 1962) and subsequently used in context of consumer credit behaviour (Frydman et al 1985). This is closely related to the customer relationship behaviour of this paper and so we can apply the model to the standard {1, 2, 3+} state space but assuming only that the stayers will stay in state 1. In terms of the notation in section two this means we assume $s_2 = s_3 = 0$ and it implies that we believe the population is divided into two groups – one who only will make one purchase with the company (the stayers) and a second to whom it is possible to sell a number of different products. It does not seem sensible to assume there are people who will make two and only two purchases with the organisation. Using the estimation procedure given by Frydman et al (1985) we found that $s_1 = 0.87$ and that the probability transition matrix for the first order time homogeneous Markov chain assumed to describe the movers is of the form

$$Q_{mover} = \begin{vmatrix} 0.99363 & 0.0632 & 0.005 \\ - & 0.9826 & 0.0174 \\ - & - & 1.000 \end{vmatrix}$$

The forecasting accuracy of this model is compared with the base model in Figure 4 by looking at the estimated cumulative purchases over a number of time periods using the holdout sample. This shows the model is quite accurate in the short term future, but the longer term forecasts are not as good as the basic model. It is perhaps surprising (and perhaps disappointing to the organisation) that the stayers comprise such a large proportion (87%) of the population. Predicting who are movers and who are stayers would allow more focussed targeting of the marketing effort. This was done by Ho et al (2004) in the case of a Markov chain describing the credit behaviour of bank customers though there one was discriminating between four groups – movers, stayers, twitchers and shakers.

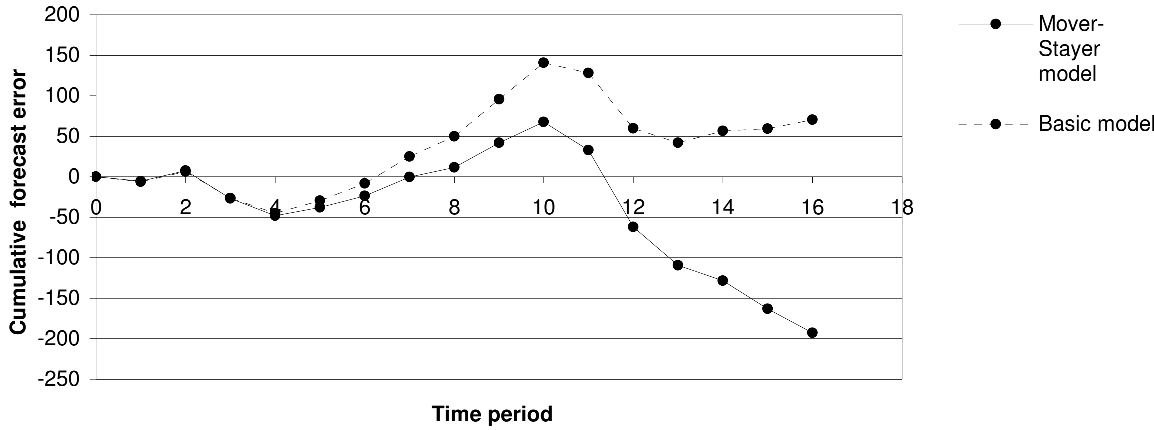

Figure 4. Cumulative forecasting error using mover-stayer model

Instead of using extra information to segment the population and then building models with the same state spaces but different transition matrices on each segment, one could include extra information in the state space itself. One would expect that the knowledge of the products that had been purchased may be useful in predicting future purchase behaviour and so we consider two models where this information is included.

The products sold by the company are segmented into four main types - A,B,C and D - consisting of investment products and various type of insurance and pension products, though for confidentiality we do not name them specifically. In the last purchase model the state is (i, X) where i =1,2 or 3+ is the cumulative number of products purchased and X is the type of the last product purchased. Thus we appear to have a Markov chain with 12 states. If however we combine the four 3+ states into one since they are all essentially absorbing states we can cut the state space down to 9 states. Concentrating on the time homogeneous first order version, we can estimate the transition probabilities in the usual way. The resultant matrix Q has the following form, where the states are ordered (1,A), (1,B),(1,C), (2,A) etc. Notice that each row will only have 5 non-zero entries corresponding to staying where it is or moving to the state with one more purchase.

$$Q_{\substack{12state\\ model}} = \begin{vmatrix} 0.993 & - & - & - & 0.007 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\ - & 0.990 & - & - & 0.001 & 0.005 & 0.003 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\ - & - & 0.988 & - & 0.000 & 0.001 & 0.011 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\ - & - & - & 0.996 & 0.001 & 0.001 & 0.001 & 0.001 & 0.000 & 0.000 & 0.000 & 0.000 \\ - & - & - & - & 0.984 & - & - & - & 0.014 & 0.001 & 0.001 & 0.001 \\ - & - & - & - & - & 0.970 & - & - & 0.004 & 0.017 & 0.009 & 0.000 \\ - & - & - & - & - & - & 0.982 & - & 0.001 & 0.001 & 0.017 & 0.000 \\ - & - & - & - & - & - & - & 0.990 & 0.005 & 0.000 & 0.001 & 0.004 \\ - & - & - & - & - & - & - & - & 1.000 & - & - & - \\ - & - & - & - & - & - & - & - & - & 1.000 & - & - \\ - & - & - & - & - & - & - & - & - & - & 1.000 & - \\ - & - & - & - & - & - & - & - & - & - & - & 1.000 \end{vmatrix}$$

What this matrix shows is there is a strong tendency to repeat buy (i.e. the second purchase is of the same type as the first). This comes from the probabilities 0.007, 0.005, 0.011 in the first three rows. The only exception is product D where the chance of any second purchase is low and likely to be of any type (shown by the values 0.996, 0.001, 0.001, 0.001 and 0.001 in the fourth row). A similar repeat purchase pattern occurs for the third purchase though the sequence B $\rightarrow$ C is quite likely and the sequence D $\rightarrow$ A is more likely than D $\rightarrow$ D. The predictive ability of this model is shown in Figure 5. It has consistently lower forecasting error than the base model which ignores the purchase type.

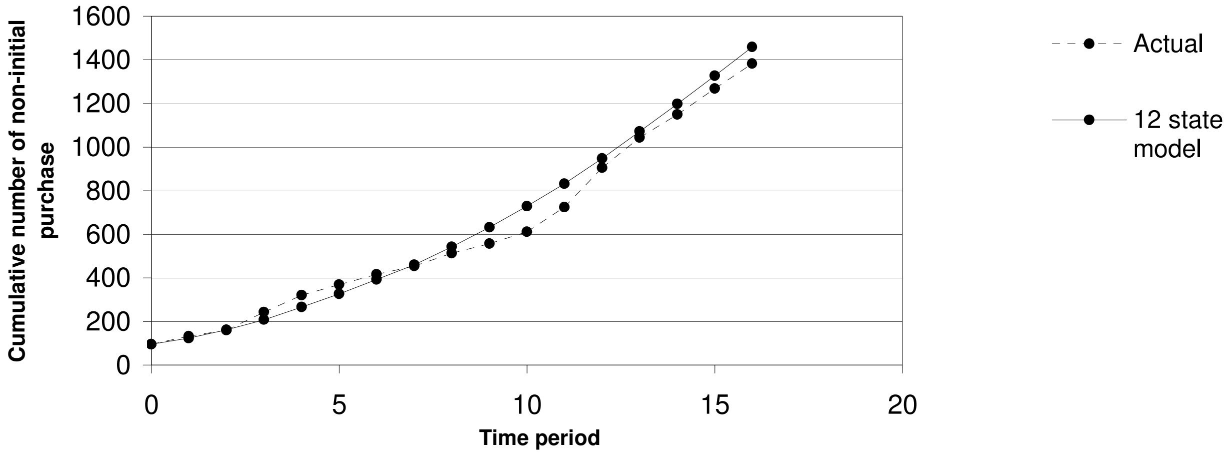

Figure 5.a Actual and 12 state model (3-month period) forecast for cumulative number of non-initial purchase

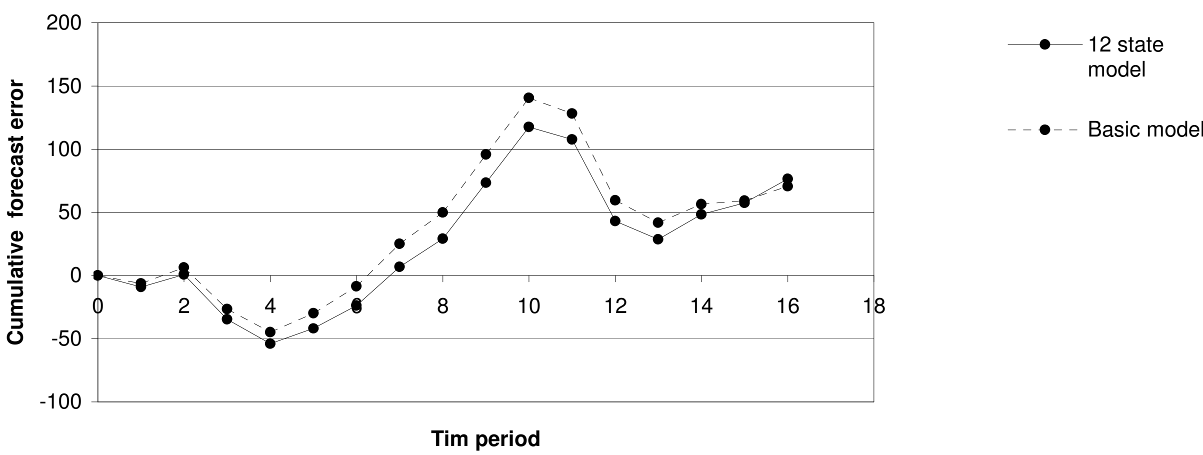

Figure 5.b Cumulative forecasting error for 12 state model

The next model, the full product type model looks at the cumulative number of purchases of each type of product separately. Thus the states are of the form (a,b,c,d) where a is the number of A purchase already made, b the number of B purchase already made and similarly for c and d. in this model it is possible for a customer to have 0 purchases of a given type and so the values of each component could be 0,1,2, etc. Even if we decide to amalgamate all second and higher purchases of a type and so only have states 0,1 and 2+ in each type this leads to a Markov chain with 80 states ((0,0,0,0) is not possible). This is partly because we are distinguishing states with up to 8 purchases but it does cause real problems with estimation since many of the states will occur very rarely even with a sample of 50,000.

One way around this is to consider the purchases of each product to be independent of one another. This is a very strong assumption but does mean we can separate out the process into four independent Markov chains each with a state space of 0,1,2,3+ etc cumulative purchases of that type. If we again go with 0,1,2,3+, we end up with 16 states in total (four chains with 4 states in each) but since only q_{00}, q_{01}, q_{02}, q_{11}, q_{12} and q_{22} are independent estimates in each chain there are only 24 parameters to estimate. This compares with almost 1000 transition probabilities to be estimates if independence is not assumed. We call this the independent type model and the simplest version of such a model assumes each of these chains is first order and stationary. The estimates from our data were as follows

$$Q(A)=\begin{vmatrix} 0.9995191 & 0.0004809 & 0.0000000 & 0.0000000 \\ - & 0.9933324 & 0.0066236 & 0.0000440 \\ - & - & 0.9859978 & 0.0140022 \\ - & - & - & 1.0000000 \end{vmatrix}$$

$$Q(B)=\begin{vmatrix} 0.9994730 & 0.0005270 & 0.0000000 & 0.0000000 \\ - & 0.9952841 & 0.0047159 & 0.0000000 \\ - & - & 0.9889706 & 0.0110294 \\ - & - & - & 1.0000000 \end{vmatrix}$$

$$Q(C)=\begin{vmatrix} 0.9993873 & 0.0006127 & 0.0000000 & 0.0000000 \\ - & 0.9894055 & 0.0105447 & 0.0000498 \\ - & - & 0.9796000 & 0.0204000 \\ - & - & - & 1.0000000 \end{vmatrix}$$

$$Q(D)=\begin{vmatrix} 0.9996907 & 0.0003093 & 0.0000000 & 0.0000000 \\ - & 0.9990572 & 0.0009428 & 0.0000000 \\ - & - & 0.9972145 & 0.0027855 \\ - & - & - & 1.0000000 \end{vmatrix}$$

We again can use this model to estimate the total number of purchases in any period by estimating the total of each type and then summing these estimates. Using the usual holdout sample, Figure 6a gives the expected number of purchases given by the independent type model and the actual number of purchases made while Figure 6b compares the error of this independent model with that of the basic model. Figure 6 suggests that the independence assumption is too strong in that the forecast errors in the cumulative number of purchases are worse than in the base case. So this approach does not seem to be particularly appealing in that it leads either to Markov chains with large state spaces or to the independent version which loses useful information concerning the purchases of the other types of products.

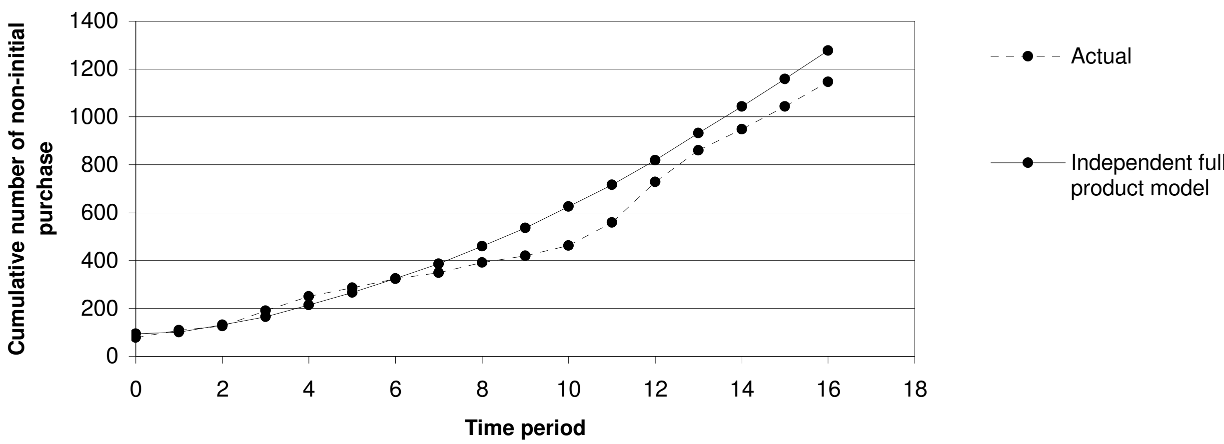

Figure 6.a Actual and independent full product model forecasts of cumulative numbers of non-initial purchase

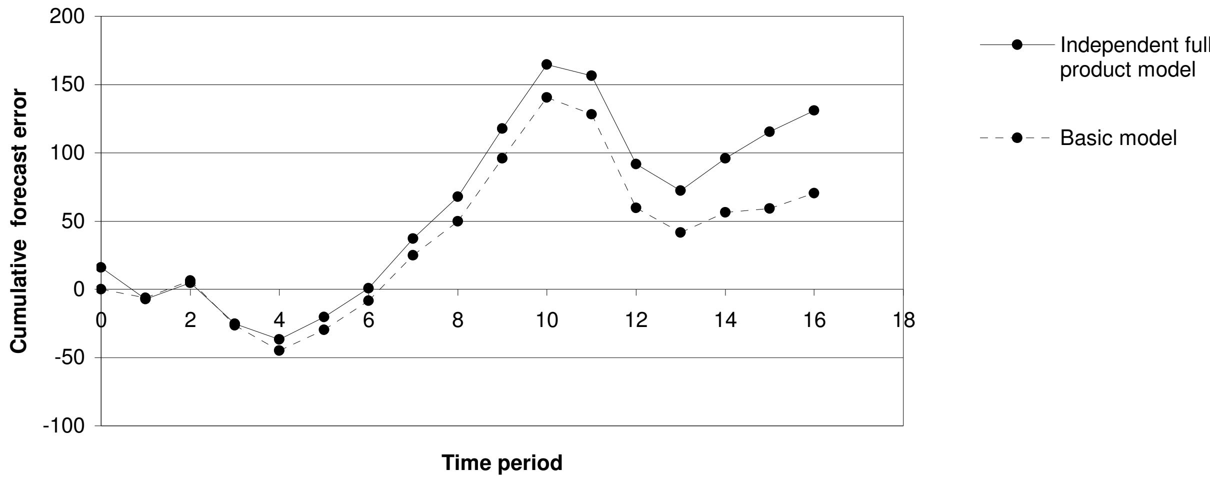

Figure 6.b Cumulative forecasting errors for independent full product model

7 Economy dependent Markov Chain Models

In Section 5, the fully non-stationary model where there are different transition probabilities for each time period, was shown to give good forecasts of the number of purchase on the hold-out sample over the time period it was built on. However this is of no use if one is trying to forecast outside the time periods on which the model was developed. If these transition probabilities can be related to other variables and one can forecast these other variables for future time periods, one could develop transition probability matrices based in these forecasts which can then be used for modelling purchase behaviour in future time periods.

Obvious candidates for these external variables are the economic indicators which relate to consumers. To this end we concentrated on the five economic variables described in section three, namely stock market returns, consumer price index, consumer confidence index, unemployment rate and the interest rate. The stock market returns impact consumer purchase behaviour in several ways. Firstly many of the investment product sold by the insurance company have returns dependent on the stock market and so look more attractive when the stock market is rising. Also rises in the stock market suggest a growing economy which gives more confidence to consumers, while for a small but significant (certainly for pension and investment products) group of consumers it increases the bonuses they are likely to get. One would expect rises in the consumer price index to mean consumers have less money available for purchasing extra products, though of course this would also depend on what is happening to wages at the same time (a wage index was tried but added no improvement to the model presented here). Rises in consumer confidence could mean that consumers are more willing to risk overstretching themselves by purchasing more products or contracting to invest more because they are more confident of sustained higher income levels. However this willingness to take risks might also mean that for insurance companies there may be a lowering of the purchasing of protection products. Initially one might expect that increases in the unemployment rate would lead to a drop in the number of new products being purchased, but there are some counter arguments. The consumers who invest in insurance company products are usually financially aware, tend to have reasonably high incomes and have a propensity to be cautious. These people are less likely to lose their jobs than the average person when unemployment rises and for some such rises will trigger a need to purchase products that protect themselves and their families. Lastly interest rate rises could affect purchase behaviour in both directions. They may trigger rises in mortgage rates and hence decrease the disposable income of some of the potential purchasers. On the other hand for those with savings but no major interest linked outgoings their disposable income will go up.

To build an economy dependent Markov chain model we will continue to use the three monthly (quarterly) transition period we chose in section three and take the value of the economic variable in a quarter to be its value at the start of the quarter. Again to keep the models as simple as possible we concentrate on building a first order Markov chain. So for each of the sixteen quarters in the sample time period we have values of the economic variables during the quarter and for each pair of states i and j the number of people who started the quarter in state i and start the next quarter in state j. Thus for state 1 we have the number of people who stay in state 1, the number who move to state 2 and the number who move to state 3+ and from this data we want to relate the probabilities of such moves to the values of the economic variables. This can be modelled as a 3-class logistic regression and the obvious ordering of the states makes this an obvious candidate for the ordered or cumulative variant of such models. In such regressions, if q_{ij} is the probability of going from state i to state j and $\mathrm{Q}_{\mathrm{ik}} = \sum_{j:j\leq k} q_{ij}$ is the probability of moving to the states k or less, then if the variables are $\mathbf{x} = (\mathrm{x}_1, x_2, .., x_n)$ the relationship is

$$\ln\left[\frac{Q_{ik}}{1-Q_{ik}}\right] = \alpha_{ik} + \sum_j \beta_{ij}\, x_j \qquad \text{for all } i,k \in S$$

This can be rewritten so that for state 1, if one defined

$$A_k = e^{\left[\alpha_{1k} + \sum_j \beta_{1j}\, x_j\right]}$$

Then $\quad q_{11} = \dfrac{1}{1+A_1^{-1}} \quad$ and $\quad q_{11} + q_{12} = \dfrac{1}{1+A_2^{-1}}$

Similarly for state 2 if

$$B_k = e^{\left[\alpha_{2k} + \sum_j \beta_{2j} x_j\right]}$$

then

$$q_{22} = \frac{1}{1 + B_2^{-1}}$$

For each variable x_j if the coefficient β_j is positive then as x_j increases the values q_{11} and $q_{11} + q_{12}$ also increase. That means there is less chance of a purchase, while if the coefficient β_j is negative then as the variable increases there is more chance of a purchase.

Using the data in the development sample we built for each of the five economic variables a single variable model connecting the transition probabilities to that variable. Four of these models were unremarkable but in the unemployment variable model, the coefficients were negative, which suggests that as unemployment rises the chance of purchase also increases. This does seem to be a real feature of the data and suggests the increase in risk aversion as unemployment increases more than compensates for the loss of income among those who become unemployed. However we felt the impact of this variable is suspect over a large time horizon and so left it out of the subsequent analysis.

If the four remaining economic variables are put in the model together the α and β coefficient estimates are given in Table 5. The figures in brackets are the standard errors in the estimates and the starred values are the estimates which are significantly non-zero at the 95% level. Looking at the effects it seems that in all cases an increase in the stock market index leads to an increase in the likelihood of purchase both for consumers who have already made 1 or 2 purchases. Similarly increase in the interest rate means both sets (those with 1 and with 2 purchases) of consumers are less likely to purchase which suggests it may be the mortgage rate effect which is dominating. Similarly the model suggests that increasing consumer confidence means increasing likelihood of purchasing and even if the impact is small it is significant. Consumer prices have a less clear impact, since neither the transitions from state 1 nor the transitions from state 2 have significant coefficients on consumer prices.

	Stock Market	Interest Rate	Consumer Price	Consumer Confidence	α_1	α_2
Coefficient in state 1	-2.087*	0.197*	0.9396	-0.053*	3.5338*	8.341*
Model q_{1k}	(0.238)	(0.034)	(3.190)	(0.011)	(0.170)	(0.285)
Coefficient in state 2	-3.004*	0.505*	-17.065	-0.085*		1.635*
Model q_{2k}	(0.777)	(0.112)	(9.947)	(0.035)		(0.551)

Table 5. Coefficients in cumulative logistic model with four economic variables

Since the impact of changes in consumer confidence is small we redo the model with this variable left out. Table 6 shows the coefficient estimation values for this simpler model. The impacts of the stock market and the interest rate are still significant for both transitions from states 1 and states 2 and they still have the same signs. So increases in the stock market encourages purchasing and increase in the bank interest rate discourages purchasing. In this model consumer prices affect consumers who have made 1 purchase in that rises in prices discourages another purchase. This suggests that inflation may affect the sense of economic well-being though inflation in itself describes money illusion which should not affect real economic behaviour. Perhaps inflation has an effect on sentiment which replaces our (weak) consumer confidence variable in Table 5.

	Stock Market	Interest Rate	Consumer Price	α_1	α_2
Coefficient in state 1	-1,807*	0.093*	7.372*	4.137*	8.945*
Model q_{1k}	(0.226)	(0.026)	(2.811)	(0.122)	(0.259)
Coefficient in state 2	-2.088*	0.328*	-3.532		2.585*
Model q_{2k}	(0.638)	(0.084)	(7.620)		(0.399)

Table 6. Coefficients in cumulative logistic model with three economic variables

The economy dependent Markov chain models we have constructed can be used to forecast total future purchases provided one is able to forecast the future values of the economic variables. Such forecasts of purchase numbers would have two sources of error – the modelling error and the error in the forecasts of the economic variables. In order to compare this model with the previous ones considered we will only consider one of these errors by comparing the actual number of purchases on the holdout sample with that predicted by this model on the same four year period as the model was built on. We assume that the forecasted value of the economic variable was the actual value that occurred i.e. perfect forecasts. It is interesting to compare the number of purchases estimated under this model compared with the actual number purchased in the hold out sample and the number using the completely non-stationary model. The latter model corresponds to getting a perfect description of the time non-homogeneity of the process by the changes in the economic variables. The results of Figure 7 show these non-stationary models do mimic the kink in actual purchases about quarter 10. Standard Markov chains, on the other hand, always lead to a smooth cumulative purchase forecast curve. The economic model is also significantly better than the base model in predicting the cumulative number of purchases for the first three years. It does not do so well in the final year though and one has to remember that in reality one will have errors from forecasting the economic variables as well. The advantage of such a model though is that it is compatible with the insurance organisations other business models, most of which include economic forecasts as part of their input.

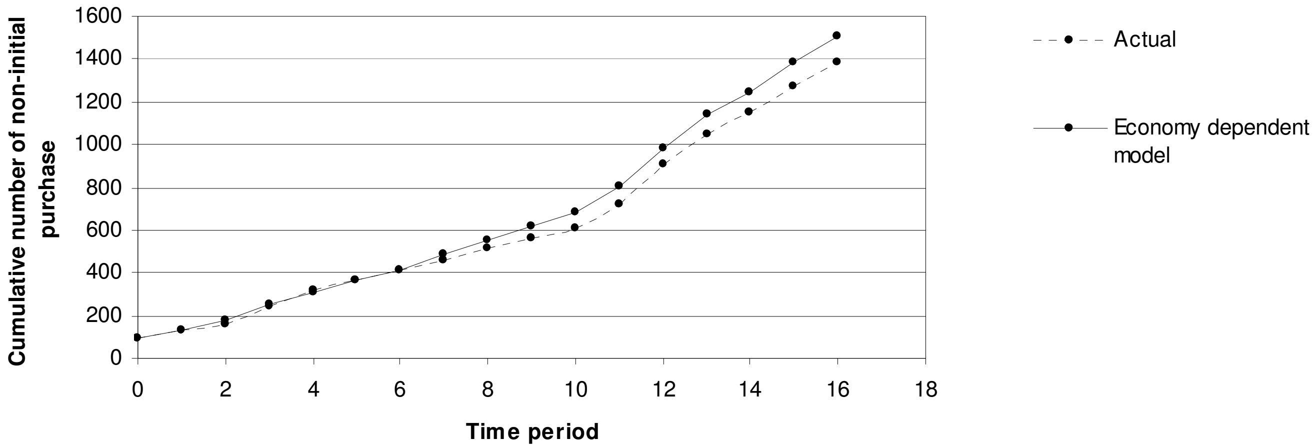

Figure 7.a Actual and economy dependent model (3-month period) forecast for cumulative number of non-initial purchase

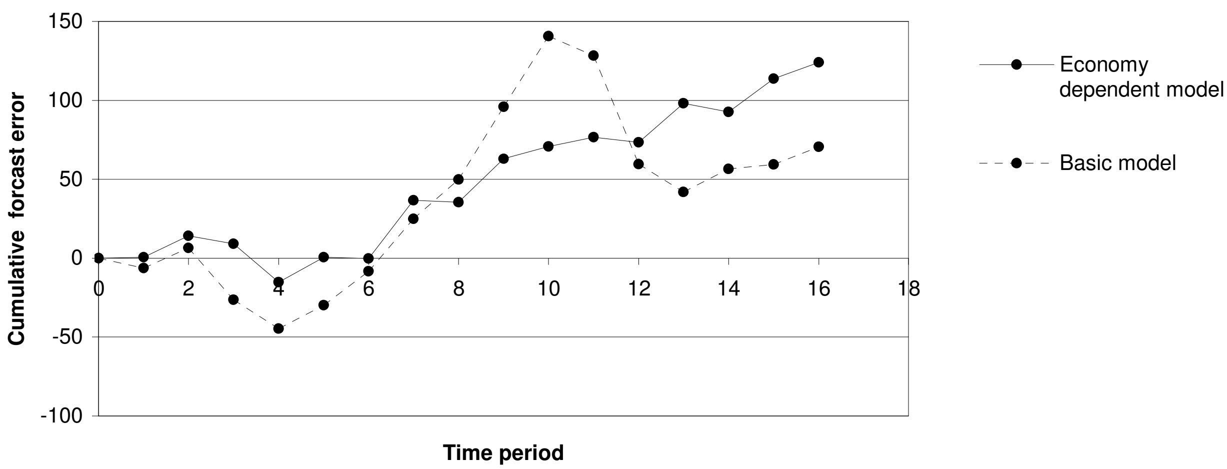

Figure 7.b Cumulative forecasting errors using economy dependent model

8 Conclusions

The paper considers how various types of Markov chain model can be used to help forecast the purchase behaviour of consumers. The models were used in a case study using purchasing behaviour for customers of a major insurance company. This led to some specificity in the modelling- for example the state spaces tend not to be very large because the number of customers making more than 3 purchases is very small- but allows one to make comparisons with what happened in reality by using holdout samples. The paper looks at how one can choose appropriate transition time periods (Here we have assumed they are all the same length, but one could increase the length of the transition period the further from the present one is modelling). We examine generalizations of the basic time homogeneous first order Markov model - the one used almost exclusively in the literature to date to model customer behaviour. The results suggest that going to a second order Markov chain could be worthwhile in statistical terms but the improvement in performance is not great. Allowing for time non-homogeneity is sensible and the seasonal model displays real forecasting improvement. The completely non-homogeneous time model is also accurate but does not lead to an operable forecasting procedure. The way to overcome this is to build an economy dependent Markov chain, and we showed how to build such models and that their forecasting accuracy is promising.

One could extend these models in many ways; by looking at second and higher order time dependent Markov chains; by allowing time dependent and second order variants of the mover stayer models and the product dependent purchase models. One could allow other forms of introducing the economic effects into the transition matrices. Ho (2002) for example uses a constrained linear regression approach to connect the transition probabilities to the economic variables in a consumer credit risk context. Nickell et al (2000) used an ordered probit model to estimate the transition probabilities of the credit rating grades of bonds in terms of the economic conditions. One could also use independent logistic regression approaches for each transition probability in the row of the transition matrix if one felt there was no ordering within the state space. However our aim was to investigate the relative effect of each of these generalizations and to build models that are sufficiently simple as to be understandable.

This paper has considered modelling the purchase behaviour of customers. This is the most important element in modelling the customers' overall behaviour and hence in building a customer lifetime value model and a customer relationship model. We have not distinguished here between products which are still in the process of being purchased and one whose purchase is complete. (in insurance and assurance one can be "buying" a product over a number of years) but this is only a matter of defining the states as the number of active products as well as the total number of purchased products. One can deal with customers who stop using a product by allowing transitions to states with a smaller number of active products as well as to ones with more active products. One can model customers who sever all connection with the company by adding an absorbing state corresponding to the "death" of the customer – in this context this could be literally true.

We feel this paper is useful in developing models for customer lifetime value and customer relationship management. It shows that Markov chains are a feasible and flexible way of developing such models, and thus make it easy to introduce the impact of external economic effects into the forecasting procedure.

9 References

Anderson T.W., Goodman L.A., (1957), Statistical Inference about Markov chains, Annals. Math. Statis. 28, 89-109.

Blumen I., Kogan M., McCarthy P.J., (1962), The Industrial Mobility of Labor as a Probability Process, Cornell University Press, Ithica, NY.

Cyert R.M., Davidson H.J., Thompson G.L., (1962), Estimation of allowance for doubtful accounts by Markov Chains, Management Science 8, 287-303.

Dwyer F.R., (1989),Customer Lifetime valuation to support marketing decision making, Journal of Direct Marketing 3, 8-15.

Feller W., (1957), An Introduction to Probability Theory and its Applications Volume I, Wiley, New York.

Frydman H, Kalberg J.G., Kao D-L, (1985),Testing the adequacy of Markov chain and mover-stayer models as representation of credit behaviour, Operations Research 33, 1203-1214..

Gronroos C. (2000), Service Management and Marketing, a customer relationship management approach, Wiley, Chichester.

Ho J., (2001), Modelling bank Customers Behaviour using Data Warehouses and Incorporating Economic indicators, Ph.D. thesis, University of Edinburgh, Edinburgh.

Ho J., Thomas L.C., Pomroy T.A., Scherer W.T., (2004), Segmentation in Markov chain consumer credit behavioural models, in Readings in Credit Scoring (ed Thomas L.C., Edelman D.B., Crook J.N.,), Oxford University Press, Oxford, pp 295-308

Iosifescu M., (1980), Finite Markov Processes and their Applications, Wiley, New York.

Kijima M., (2003), Stochastic processes with applications to finance, Chapman and Hall, Boca Raton.,

Nickell P., Perraudin W., Varoli S.,(2001), Stability of ratings transitions, Working paper 132, Bank of England, London.

Pfeifer P.E., Carraway R.L., (2000), Modelling customer relationships as Markov chains, Journal of Interactive Marketing 14, 43-55.

Schniederjans M.J., Loch K.D., (1994), An aid for strategic marketing in the banking industry: a Markov analysis, Computers and Operations Research 21, 281-287.

Shepard, David Associates, (1995), The New Direct Marketing, Irwin, Burr Ridge, Illinois.

Till R.J., Hand D.J., (2003), Behavioural models of credit card usage, Working Paper TR-03-04, Department of Statistics, Imperial College, London.

Trench M.S., Pederson, Lau E.T., Ma L., Wang H., Nair S.K., (2002), Credit Line and Price management for Bank One Credit Cards, Working Paper, Bank One Card Services, Wilmington, Delaware.