Dynamic Segmentation using Markov-Switching Model

Takuya SATOMURA

Discussion Paper 01-09

April 2001

この研究は「大学院経済学研究科・経済学部記念事業」
基金より援助を受けた、記して感謝する。

Graduate School of Economics and
Osaka School of International Public Policy (OSIPP)
Osaka University, Toyonaka, Osaka 560-0043, Japan

Dynamic Segmentation using Markov-Switching Model

Takuya SATOMURA

Graduate School of Economics, Osaka University

1-7 Machikaneyama, Toyonaka, Osaka, 560-0043, Japan

telephone +81-6-6850-5233 fax +81-6-6850-5277

e-mail:satomura@econ.osaka-u.ac.jp

April 9, 2001

abstract

In this paper, we propose a new methodology of modeling dynamic segmentation. A probability that one belongs to a segment is defined as a markov process, and a probability that one chooses a brand is defined as a multinomial logit model. The estimation of such model has been difficult because of the complicated calculation of log likelihood. Using *Markov-switching model*, we can estimate the model parameters and latent states (=segments) at each purchase occasion.

In the empirical study of using scanner panel data, We have estimated an instability of segment in the instant coffee category. We assumed that there are two preference structures, one is under the influence of loyalty and the other is not. The empirical result showed that there are some people that change their own segment within data period.

keywords: dynamic segmentation, Markov-switching model, state space model, choice models

Introduction

In the marketing literature, the stability of a segment is one of the most important criteria. Most probabilistic choice models in the marketing area assume the stability of preference and/or choice processes. However, if the preference and/or choice processes change over time, we will fail to identify the segment using a model based on the assumption of stability. Modeling the instability of segment is necessary for the description of the change of structure.

There are two major approaches to take into account the instability of segment. One approach is *manifest change*, and the other is *latent change*(Wedel and Kamakura 1998 p.159). In the manifest change, the segment membership is stable, but changes may occur in the preference or choice structure of customers in a segment over time. In the latent change, the preference structure of segments is stable, but changes may occur in segment size and/or the segment membership of consumers over time.

In this paper, we will take into account not only that individuals differ in their preferences but also that they change a segment which they belong to. So, latent change is proper for our application.

In a latent class model, mixture components are used as a prior of the segment size, and assumed that mixture components may change over a time. There are two main approaches in this class of models, based on different reparameterizations of the priors for the mixture components(Wedel and Kamakura 1998 p.168). One is *concomitant variable model*, and the other is *latent Markov model*.

The concomitant variable models assume that latent class probabilities depend on observable variables(Wedel and Kamakura 1998 p.168; Yang and Allenby 2000).

These models need observable variables to define segment change.

The latent Markov models assume that latent class probabilities depend on markov-switching, so we do not need concomitant variables to define segment change (Poulsen 1990; Böckenholt and Langeheine 1996; Ramaswamy 1997). However, if the number of finite time period increased, multiple integral is necessary to obtain the likelihood of the latent Markov model. Those latent Markov models are not sufficient to express the change if the number of fixed time period increases.

To solve this problems, We adopt a *non-Gaussian state-space modeling of nonstationary time series* (Kitagawa 1987). When the state-space is represented as Markov process in the state-space model, it is called *Markov-switching model*(Hamilton 1989). We present our proposed model at the next section.

1 Model

1.1 Modelling Markov-Switching process

Suppose that the hth household in the panel ($h = 1, 2, \ldots, N$) is faced with a choice of M brands at the tth purchase occasion ($t = 1, 2, \ldots, T_h$). Following the finite mixture formulation of Kamakura and Russell(1989), we also assume that there exist S market segments. Segment $s(= 1, 2, \ldots, S)$ contains households who have relatively similar preference and responses to marketing mix variables.

Let S_{ht} denote the random variable that takes a value s, and $S_{ht} = s$ when hth household belongs to segment s at purchase occasion t. Let $y_{ht} = m$ denote the event that household h chooses brand m at purchase occasion t , and $\psi_{ht} = (y_{h1}, \ldots, y_{ht})$.

We assume that evolution of S_{ht} depends upon $S_{h,t-1}$, in which case the process of S_t is named as *first order Markov-switching process.*

Denote the element of transition matrix from state i to state$j$ by p_{ij}, and let

$$\begin{aligned} & Pr(S_{ht}=j|S_{h,t-1}=i,\psi_{h,t-1}) = Pr(S_{ht}=j|S_{h,t-1}=i) = p_{ij} \qquad (1) \\ & 0 \le p_{ij} \le 1, \qquad \sum_{j=1}^{S} p_{ij} = 1. \end{aligned}$$

1.2 Modeling Brand choice

We assume that choice of a brand in the segment s is governed by a multinomial logit model. Let $\boldsymbol{x}_{htm}$ denote the vector of attributes. Then the conditional probability that brand m is chosen, given $S_{ht}=s$, is specified as

$$Pr(y_{ht}=m|S_{ht}=s) = \frac{\exp(\boldsymbol{x}'_{htm}\boldsymbol{\beta}_s)}{\sum_{i=1}^{M}\exp(\boldsymbol{x}'_{htm}\boldsymbol{\beta}_s)}. \qquad (2)$$

where $\boldsymbol{\beta}_s$ is a parameter vector in segment s.

The unconditional probability is given by the weighted sum of segment probabilities.

$$Pr(y_{ht}=m|\psi_{h,t-1}) = \sum_{s=1}^{S} Pr(S_{ht}=s|\psi_{h,t-1})Pr(y_{ht}=m|S_{ht}=s). \qquad (3)$$

1.3 Estimation

For parameter and state estimation, we need to evaluate $Pr(S_{ht}|\psi_{ht})$, the conditional probability of $S_{ht}=s$ given observations $\psi_{ht}(=y_{h1},\dots,y_{ht})$. Recursive formulas for obtaining the *One Step Ahead Prediction*, *Filtering*, *Smoothing* are as follows:

One Step Ahead Prediction:

Calculate the $Pr(S_{ht}=j|\psi_{h,t-1})$ using $\psi_{h,t-1}$ and equation(1),

$$\begin{aligned} Pr(S_{ht}=j|\psi_{h,t-1}) &= \sum_{i=1}^{S} Pr(S_{ht}=j, S_{h,t-1}=i|\psi_{h,t-1}). \\ &= \sum_{i=1}^{S} Pr(S_{ht}=j|S_{h,t-1}=i,\psi_{h,t-1})Pr(S_{h,t-1}=i|\psi_{h,t-1}) \\ &= \sum_{i=1}^{S} p_{ij} Pr(S_{h,t-1}=i|\psi_{h,t-1}). \qquad (4) \end{aligned}$$

Filtering:

Update the $Pr(S_{ht}=j|\psi_{ht})$ using Bayes' law,

$$\begin{aligned} Pr(S_{ht}=j|\psi_{ht}) &= Pr(S_{ht}=j|y_{ht},\psi_{h,t-1}) \\ &= \frac{p(y_{ht}|S_{ht}=j,\psi_{h,t-1})Pr(S_{ht}=j|\psi_{h,t-1})}{Pr(y_{ht}|\psi_{h,t-1})}. \end{aligned} \tag{5}$$

where

$$p(y_{ht}|\psi_{h,t-1}) = \sum_{j=1}^{S} p(y_{ht}|S_{ht}=j,\psi_{h,t-1})Pr(S_{ht}=j|\psi_{h,t-1}). \tag{6}$$

Smoothing:

After parameters of the model are estimated, we can make inference on S_{ht} using all information (=ψ_{hT_h}) in the sample. $Pr(S_{ht}=j|\psi_{hT_h})$ is the smoothed probability. Oppositely, $Pr(S_{ht}=j|\psi_{ht})$ is the filtered probability.

Calculate the joint probability of $S_{ht}=j$ and $S_{h,t+1}=k$ based on ψ_{hT_h}:

$$\begin{aligned} & Pr(S_{ht}=j,S_{h,t+1}=k|\psi_{hT_h}) \\ = \;& Pr(S_{h,t+1}=k|\psi_{hT_h})Pr(S_{h,t}=j|S_{h,t+1}=k,\psi_{hT_h}) \\ = \;& Pr(S_{h,t+1}=k|\psi_{hT_h})Pr(S_{h,t}=j|S_{h,t+1}=k,\psi_{ht}) \\ = \;& \frac{Pr(S_{h,t+1}=k|\psi_{hT_h})Pr(S_{h,t}=j,S_{h,t+1}=k|\psi_{ht})}{Pr(S_{h,t+1}=k|\psi_{ht})} \\ = \;& \frac{Pr(S_{h,t+1}=k|\psi_{hT_h})Pr(S_{h,t}=j|\psi_{ht})Pr(S_{h,t+1}=k|S_{ht}=j)}{Pr(S_{h,t+1}=k|\psi_{ht})}. \end{aligned} \tag{7}$$

and $Pr(S_{ht}=j|\psi_{hT_h})$ is

$$Pr(S_{ht}=j|\psi_{hT_h}) = \sum_{k=1}^{S} Pr(S_{ht}=j,S_{h,t+1}=k|\psi_{hT_h}). \tag{8}$$

Begin with $p(S_{h,T_h-1}=j|\psi_{h,T_h})$, we can calculate $p(S_{h,T_h-2}=j|\psi_{h,T_h}),\ldots,p(S_{h,1}=$

$j|\psi_{h,T_h})$. The validity of going from the second line to the third line of equation (7) is in *Appendix:A*

1.4 Model Identification

The log likelihood function for h is given by

$$\ln L_h = \sum_{t=1}^{T_h} \ln p(y_{ht}|\psi_{h,t-1}). \tag{9}$$

And the log likelihood function for all households is given by

$$\ln L = \sum_{h=1}^{H} \ln L_h. \tag{10}$$

We can Estimate the parameter by maximizing the $\ln L$ numerically. In this approach, Akaike's information criterion(AIC) is

$$AIC = -2max \ln L + 2(number\ of\ free\ parameters). \tag{11}$$

and other infromation criterion are obtained in the same way as AIC.

1.5 Advantage of Using Markov-Switching Model

If we could not use the Markov-switching Model, the log likelihood of household h is

$$\ln L_h = \ln\{\sum_{S_{h1}} \sum_{S_{h2}} \cdots \sum_{S_{hT_h}} p(S_{h1})p(y_{h1}|S_{h1}) \prod_{t=2}^{T_h} p(S_{ht}|S_{h,t-1})p(y_{ht}|S_{ht})\}. \tag{12}$$

Compared with (9), (12) is so difficult to find the parameters which maximize log likelihood. Markov-switching model has advantage of simplicity at the calculation of the log likelihood.

2 APPLICATION

In this section we estimate the model using scanner panel data set[1].

2.1 Data

The data analyzed here consist of 52 weeks scanner panel records of households in the Tokyo metropolitan area, during the year 1993. The number of 97 households who purchased instant coffee between 12 and 24 times during this period was selected.

To keep the model estimation manageable, we chose four top-selling brands(*BrandA*–*BrandD*), and the rest was gathered to one brand(*Brand E*[2]).

Brand	Market share	Average Discount Rate
A	0.286	0.794
B	0.171	0.777
C	0.133	0.747
D	0.078	0.835
E	0.331	0.734

The purpose of this analysis is to know whether people change their preference structure. The effect of loyalty(last purchased brand) to the utility may differ over the person. Will s/he be also stable all over the time? If not, when did s/he change preference structure? This information will be very important when we plan the relationship marketing(ex. electronic couponing) .

2.2 Model Variables

We begin by specifying the functional form of the choice model. The deterministic component of utility for brand m at time t, given that household h is a member of segment s, is

$$\boldsymbol{x}'_{htm}\boldsymbol{\beta}_s = \beta_{s,price}PRICE_{mt} + \beta_{s,display}DISPLAY_{mt} + \beta_{s,lbp}LBP_{mt} + \beta_{s,int,m}. \quad (13)$$

where:

$$\begin{aligned}
PRICE_{mt} &= \text{(actual shelf price of brand} m \text{ at time } t) \\
&\quad \div\text{(highest shelf price of brand} m \text{ during the period).} \\
DISPLAY_{mt} &= 1 \text{ if brand} m \text{ was displayed at time } t \text{ for } h \text{ ,} \\
&\quad \text{and 0 otherwise.} \\
LBP^h_{mt} &= 1 \text{ if brand } m \text{ was purchased by household } h \\
&\quad \text{at time } t-1\text{, and 0 otherwise.} \\
\beta_{s,price}, \beta_{s,display}, \beta_{s,lbp} &= \text{parameters to be estimated.} \\
\beta_{s,int,m} &= \text{constant for brand} m\text{, segment } s\text{, to be estimated.}
\end{aligned}$$

In this application, our interest is in the dynamics of the influence of loyalty to the utility. So we consider two segments, $Segment1$ is under the influence of loyalty ($\beta_{1,lbp} \neq 0$, to be estimated) and $Segment2$ is not (set $\beta_{2,lbp} = 0$ a priori).

2.3 Steady-State Probability

To start the Filtering, we have to define the start probability S_{h0}. We employ the steady-state probability as S_{h0} (Kim and Nelson 1999 p.66). Derivation of Steady-State Probability is stated in *Appendix B*. When Segment Size was determined to be two,

$$Pr(S_{h0} = 1) = \frac{1 - p_{22}}{2 - p_{11} - p_{22}}. \quad (14)$$

$$Pr(S_{h0} = 2) = \frac{1 - p_{11}}{2 - p_{11} - p_{22}}. \quad (15)$$

2.4 Estimation Results

Model Comparison

Before the estimation, we set three comparison models.

Model 1: Logit model with Markov-switching

This model assumes that there are two segments and household will switch the segment.

$$L_h = \prod_{t=1}^{T}[\sum_{j=1}^{S} Pr(S_{ht} = j|\psi_{h,t-1})p(y_{ht}|S_{ht} = j)]. \tag{16}$$

Model 2: Logit model with latent segment

This model assumes that there are two segments and any household does not switch segment.

$$L_h = \sum_{j=1}^{S}[\pi_j \prod_{t=1}^{T} p(y_{ht}|S_{ht} = j)]. \tag{17}$$

where π_j is size of segment j.

Model 3: Normal logit model

This model assumes that there is only one segment.

$$L_h = \prod_{t=1}^{T}[p(y_{ht})]. \tag{18}$$

The information criteria of the models are shown in Table 1. $\boldsymbol{AIC}$, $\boldsymbol{BIC}$ and $\boldsymbol{CAIC}$ are smallest at the Model1. We can say that Model1 is fitest for the data.

Feature of the segment

The estimated parameters of the Model1 are shown in Table 2. We have set $\beta_{2,lbp} = 0$ a priori, and $\beta_{1,lbp}$ is estimated positive. The influence of loyalty is positive in $Segment1$. Comparison of price parameters and display parameters, $\beta_{2,price} > \beta_{1,price}$ and $\beta_{2,display} > \beta_{1,display}$, so we can also say that $Segmen2$ is a promotion sensitive segment. As shown in Table 2, switching probability p_{11} and p_{22}, which is recurrent probability, are close to 1. The switching probability between $Segment1$ and $Segment2$ is very small.

Dynamics of Segmentation

The example of dynamics of segmentation are shown in Fig 1 and Fig 2). The household#1 suddenly changes segment at purchase occasion 9. On the other hand, the household#2 is stable over the time. As we intended, the dynamics of segmentation was estimated.

3 Discussion

In this study, we demonstrated a methodology for modeling the latent change between the segments in a short period. A probability that one belongs to a segment is defined as a markov process, and a probability that one chooses a brand is defined as a multinomial logit model. The estimation of such model has been difficult because of the complicated calculation of log likelihood. Using *Markov-switching model*, we can estimate the model parameters and latent states at each purchase occasion. In the empirical study, We have estimated an instability of segment in the instant coffee category. We assumed that there are two preference structures, one is under the influence of loyalty and the other is not. The empirical result showed that there are some people that change their own segment within data period.

The main limitation of our model is that the parameters of transition probability are dependent on an unobservable state variable (S_{ht}), that is an outcome of an unobservable Markov process. This means that inference of S_{hT} is based on a conditional distribution, not on a joint distribution. Another limitation is that the likelihood function of the parameters is still not presented in simple form. MCMC (Markov chain Monte Carlo) approach will enable us to solve these problems. Albert and Chib (1993) have made MCMC analysis of Markov-switching model.

Application of our model to another marketing area is also an interesting. We have not investigated why s/he switches her/his segment. If we use the concomitant variable(Kamakura, Wedel and Agrawal 1994) to the Markov-switching probability, we will be able to account for this problem. This analysis is also important from managerial perspective. The future research is expected.

Appendix A

The validity of going from the second line to the third line of equation (7) is given by Kim and Nelson(1988 p.68).

Define $\tilde{f}_{h,t+1,T_h} = (y_{h,t+1}, y_{h,t+2}, \ldots, y_{hT_h})'$, for $T_h > t$. That is, $\tilde{f}_{h,t+1,T_h}$ is the vector of observations from date $t+1$ to T_h. Then we have

$$\begin{aligned} & Pr(S_{ht} = j | S_{h,t+1} = k, \psi_{hT_h}) \\ = \;& Pr(S_{ht} = j | S_{h,t+1} = k, \tilde{f}_{h,t+1,T_h}, \psi_{ht}) \\ = \;& \frac{g(S_{ht} = j, \tilde{f}_{h,t+1,T_h} | S_{t+1} = k, \psi_{ht})}{g(\tilde{f}_{h,t+1,T_h} | S_{t+1} = k, \psi_{ht})} \end{aligned}$$

$$= \frac{Pr(S_{ht} = j | S_{h,t+1} = k, \psi_{ht}) g(\tilde{f}_{h,t+1,T_h} | S_{h,t+1} = k, S_{ht} = j, \psi_{ht})}{g(\tilde{f}_{h,t+1,T_h} | S_{h,t+1} = k, \psi_{ht})}$$

$$= Pr(S_{ht} = j | S_{h,t+1} = k, \psi_{ht}) \quad (19)$$

The above holds as $g(\tilde{f}_{h,t+1,T_h} | S_{h,t+1} = k, S_{ht} = j, \psi_{ht}) = g(\tilde{f}_{h,t+1,T_h} | S_{h,t+1} = k, \psi_{ht})$, which suggests that if $S_{h,t+1}$ were somehow known, then $y_{h,t+1}$ would contain no information about S_{ht} beyond that contained in $S_{h,t+1}$ and $\psi_{h,t}$.

Appendix B

Derivation of Steady-State Probability Used to Start the Filter is below(Kim and Nelson 1999 p.70).

$$P^* = \begin{bmatrix} p_{11} & p_{21} & \cdots & p_{S1} \\ p_{12} & p_{22} & \cdots & p_{S2} \\ \vdots & \vdots & \ddots & \vdots \\ p_{1S} & p_{2S} & \cdots & p_{SS} \end{bmatrix} \quad (20)$$

where $i'_S P^* = i'_S$ with $i_S = [1 \; 1 \; \ldots 1]'$. If we let π_t be a vector of $S \times 1$ steady-state probabilities, we have

$$\pi_t = \begin{bmatrix} Pr(S_t = 1) \\ Pr(S_t = 2) \\ \cdots \\ Pr(S_t = S) \end{bmatrix} = \begin{bmatrix} \pi_{1t} \\ \pi_{2t} \\ \cdots \\ \pi_{St} \end{bmatrix} \quad (21)$$

$$i'_S \pi_t = 1. \quad (22)$$

Then according to the definition of steady state probabilities, we have $\pi_{t+1} = P_* \pi_t$

and $\pi_{t+1} = \pi_t$, and thus

$$\pi_t = P^* \pi_t \Longrightarrow (I_S - P^*)\pi_t = 0_S. \tag{23}$$

where 0_S is an $S \times 1$ matrix of zeros. Combining equations(22)and(23) ,we have

$$\begin{bmatrix} I_S - P^* \\ i'_S \end{bmatrix} \pi_t = \begin{bmatrix} 0_S \\ 1 \end{bmatrix}, \text{ or } A\pi_t = \begin{bmatrix} 0_S \\ 1 \end{bmatrix}. \tag{24}$$

Multiply both sides of the above equation by $(A'A)^{-1}A'$. Then,

$$\pi_t = (A'A)^{-1}A' \begin{bmatrix} 0_S \\ 1 \end{bmatrix} \tag{25}$$

That is, the matrix of steady-state probabilities, π_t, is the last column of the matrix $(A'A)^{-1}A'$.

Notes

1. Scanner panel data was offered from one of the Japanese supermarket. The name of supermarket is secret because of the duty to protect privileged information. This data set is same as Moriguchi and Mori(1995).

2. The Value of Price and Display of *Brand E* was computed as following.

- If one choose one of *BrandE*, the price and the promotion of *Brand E* is this chosen Brand.
- Otherwise, the price is set to the largest discount rate and the promotion is set to 1 if at least one of *Brand E* is promoted.

References

Albert, J.H. and S. Chib (1993),"Bayesian Analysis of Binary and Polychotomous Response Data," Journal of the American Statistical Association, 88,669-679.

Böckenholt,U. and R. Langeheine (1996),"Latent Change in Recurrent Choice Data," Psychometrika, 61,285-301.

Hamilton, J.D. (1989),"A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle." Econometrica,57,357-384

Kamakura, W.A. and G.J. Russell (1989),"A Probabilistic Choice Model for Market Segmentaion and Elasticity Structure," Journal of Marketing Research, 26,379-90.

Kamakura, W.A., M. Wedel and J. Agrawal (1994),"Concomitant Variable Latent Class Models for Conjoint Analysis," International Journal of Research in Marketing, 11,451-64.

Kim, Chang-Jin, and C. R. Nelson (1999), State-Space Models with Regime Switching, The MIT Press.

Kitagawa, G. (1987),"Non-Gaussian state space modeling of nonstationary time series,(with discussion)", Journal of American statistical Association, 82,1032-1063.

Moriguchi, T. and M. Mori (1995),"A Brand Choice Model Considering Evoked Set(in Japanese)", Journal of Marketing Science (published by JIMS), 26,88-98.

Poulsen, C.S. (1990),“Mixed Markov and Latent Markov Modeling Applied to Brand Choice Behavior,” Intern. J. of Research in Marketing, 7,5-19.

Ramaswamy, V (1997),“Evolutionary preference segmentation with panel survey data: An application to new products,” Intern. J. of Research in Marketing, 14,57-80.

Yang, S. and G.M. Allenby (2000),“A Model for Observation, Structural, and Household Heterogeneity in Panel Data,” Marketing Letters, 11,137-149.

Wedel, M. and W.A.Kamakura (1998), Market Segmentation, Kluwer Academic Publishers.

Table 1: Model Comparison

	Model 1	Model 2	Model 3
LL	−1268.54	−1352.09	−1387.24
AIC	2567.08	2732.18	2788.48
BIC	2643.68	2803.68	2824.23
$CAIC$	2643.70	2803.69	2824.23

Table 2: Estimation Results

$Segment1$		$Segment2$	
β_{1price}	−2.540	β_{2price}	−12.206
	(−4.45)[a]		(−8.25)
$\beta_{1display}$	0.467	$\beta_{2display}$	0.984
	(3.17)		(6.38)
β_{1lbp}	1.977	β_{2lbp}	0[b]
	(18.49)		
β_{1int1}	−0.346	β_{2bint1}	1.425
	(−2.12)		(6.33)
β_{1int2}	−0.037	β_{2int2}	−1.261
	(−0.28)		(−3.81)
β_{1int3}	−1.378	β_{2int3}	0.291
	(−5.58)		(1.64)
β_{1int4}	−0.694	β_{2int4}	−1.112
	(−3.98)		(−3.59)
p_{11}	0.961	p_{22}	0.954
	(10.35)		(10.00)

[a] Asymptotic t-statistics in parenthese.

[b] Set to 0 a priori.

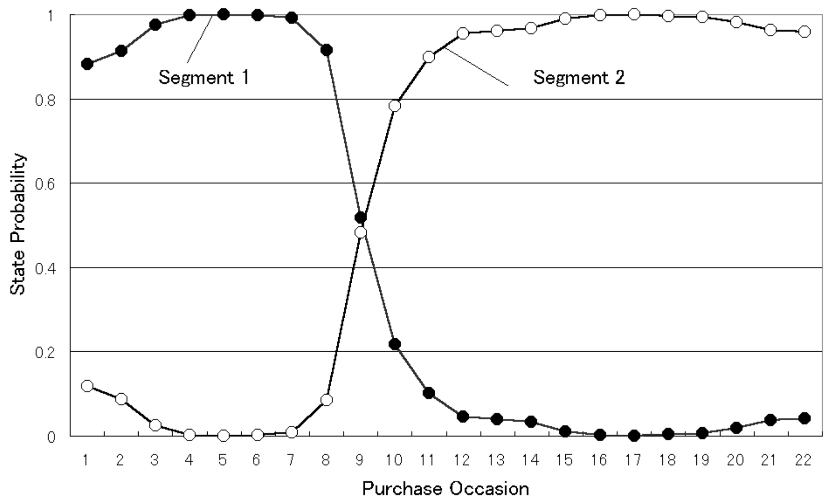

Figure 1: The dynamics of state probability in household#1

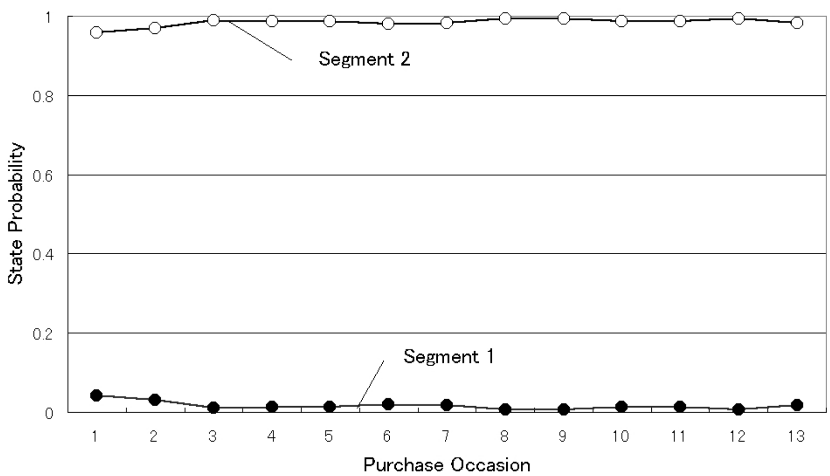

Figure 2: The dynamics of state probability in household#2