Marketing Letters 15:1, 37–60, 2004

Statistical Inference Using Stochastic Switching Models for the Discrimination of Unobserved Display Promotion from POS Data

TADAHIKO SATO* t.sato@dei.or.jp
Department of Statistical Science, The Graduate University for Advanced Studies, Distribution Economics Institute, TOC Bldg., Nishi-Gotanda Shinagawa-ku, Tokyo 141-0031, Japan

TOMOYUKI HIGUCHI higuchi@ism.ac.jp
Prediction and Knowledge Discovery Center, The Institute of Statistical Mathematics, 4-6-7 Minami-Azabu, Minato-ku, Tokyo 106-8569, Japan

GENSHIRO KITAGAWA kitagawa@ism.ac.jp
The Institute of Statistical Mathematics, 4-6-7 Minami-Azabu, Minato-ku, Tokyo 106-8569, Japan

Abstract

The execution of price and/or display promotion has a significant effect on the sales of a brand sold in a supermarket. Information on price and/or sales is available from POS data. However, unless an investigator collects information on the execution of display promotions from every retail store, such information is unavailable. This paper presents a method of identifying whether display promotion has been executed without having to visit individual stores. We treat the execution/non-execution of a display promotion as a state variable. An unknown stationary probability matrix is assumed to describe the probability of a transition between states. Each state is characterized by a different stationary time series model with unknown parameters. The objective of the analysis is to identify the model and to assign a probability model for each state at each time instant. Finally, we provide a high precision estimator of a past execution/non-execution of a display promotion based on the proposed model.

Keywords: display promotion, POS data, Markov switching model, non-Gaussian filter/smoothing

The paper begins by describing Japanese consumer purchase behavior. The shopper for a Japanese household usually visits the supermarket about 3 times per week. A supermarket in Japan corresponds to a combined grocery and drug store in the United States. A Japanese homemaker has to buy many items to satisfy the different members of the family because, for example, family members have different tastes in food. Homemakers visit the supermarket frequently and purchase many items. As a result, the non-planned purchase rate in a store is very high, about 70% (Tajima, 1989). The non-planned purchase rate is the percentage of purchased goods that were not decided before visiting the store. Moreover, in Bucklin and Lattin (1991), the extensive documentation of planned versus non-planned purchasing and its heterogeneity across consumers provided both logical and empirical support for their distinction between planned and opportunistic shopping

* Corresponding author.

modes. For effective and efficient sales, marketing activities in a store are very important, for example temporary price cut, display, Point Of Purchase (POP). To attain effective and efficient sales, suppliers need to measure the effect of those marketing activities from data.

Price promotion has significant effects on increasing sales volume (see for example (Neslin, 2002)). Information on price promotion can be acquired easily from POS data. POS data contains information on the number of units, sales and visitors of a store. This enables suppliers to decide their price strategy based on the effect of price promotion. Although suppliers are aware that execution of a display promotion in a store causes an increase in sales, suppliers usually cannot acquire information on the display promotion at each store. This means that they cannot measure the effect of execution of a display promotion. If a supplier wants to collect information on the execution of display promotion data by conducting field research, in the case of Japan, they have to pay about 4 million yen per year per store. Although this amount is an approximation, it is obviously too expensive to carry this out in many stores. As a result, effective and efficient sales with display promotion have not been realized at most stores. Therefore, if the execution/non-execution of display promotion could be inferred from POS data, this would be useful for suppliers to enhance the effectiveness and efficiency of their marketing activity.

Since temporary price cuts and other marketing promotions are important, marketing manager and academics have conducted many studies from various perspectives since the 1970's. Blattberg et al. (1981), Gupta (1988), and Neslin et al. (1985) found evidence that promotions are associated with *purchase acceleration* in terms of an increase in quantity purchased and decreased *interpurchase time.* Researchers studying the brand choice decision – for example, Guadagni and Little (1983) and Gupta (1988) – found that promotions are associated with *brand switching.* Montgomery (1971) and Webster (1965) found that promotion-prone households were associated with lower levels of *brand loyalty.* While these researches used household level scanner panel data or field research data, an alternative approach using store level POS data exits. Blattberg and Wisniewski (1987) developed a statistical model for measuring the effect of sales promotions. The model incorporates many variables for promotion activities. Another model has been developed by Wittink et al. (1987) using store level POS data and promotional data. The promotion data used in these researches, was obtained by field research in a limited number of stores. However, price strategies and display promotions differ according to variables such as geographical location, demographic features of consumers, and so forth. As a result, consumer decisions, which are made in response to these strategies, differ. For this reason, suppliers want to obtain POS data and promotional data for individual stores to enable them to promote efficient and effective sales. However, while store level POS data is easily obtained, promotional data cannot be obtained unless retailers, manufacturers or third-parties conduct an actual marketing survey on every store.

To meet this need, we constructed a model for discriminating between the execution and non-execution of display promotion from POS data without supervised data (data obtained by an in-store marketing survey). In considering the problem of identifying whether a display promotion has been conducted, we treated the execution/non-execution of a display promotion as an unobserved state variable. First, we built models in an attempt to estimate state probabilities of the execution/non-execution of a display promotion from POS data, and then determined how to discriminate the execution/non-execution of display promotion based on these state probabilities. Our estimation procedure can be regarded as unsupervised learning. The model can be formulated within the framework of a general state space model. A general state space model is constructed by two stochastic equations: a system model, and an observation model (Kitagawa and Gersch, 1996). This general state space model includes a wide variety of important time series models. Examples include the linear state space model with non-Gaussian white noise processes, the nonlinear model such as the Chaos process, the discrete process, and so forth (Higuchi and Kitagawa, 2000). For state estimation in the general state space model, we need to evaluate conditional distributions, which are in general non-Gaussian. Therefore, we need to employ non-Gaussian filtering and smoothing algorithms (Kitagawa, 1987) to evaluate conditional distributions. A statistical problem in this study is how to identify the model and assign a probability for each state at each time instant. The proposed procedure is evaluated by an observed display promotion data, i.e. supervised data, for model validation.

Generalization of the preceding time series model to include the possibility of transition changes occurring over time has been achieved by allowing changes in the error covariance (Harrison and Stevens, 1976; Gordon and Smith, 1988, 1990) or by assuming mixture distributions in the observation errors (Pena and Guttman, 1988). Changes can also be modeled in the classical regression case by allowing switches in the design matrices, as in Quandt (1972). Switching via a stationary Markov chain with independent observations has been developed by Goldfeld and Quandt (1973). Markov Switching for dependent data has been applied by Hamilton (1989) to detect changes between positive and negative growth periods in the economy.

Our paper consists of 4 sections. In the first section, Markov switching models are presented for discrimination between execution and non-execution of display promotion. Furthermore, for the state estimation, the non-Gaussian filtering and smoothing algorithms are briefly described. Our proposed model is evaluated in Section 2 with its application to actual POS data. Section 3 is devoted to the discussion on implication. The concluding remarks are given in Section 4.

1. Markov Switching Model

In this section, we introduce the general state space model, which generalizes linear Gaussian state space modeling to deal with non-Gaussian system noise and a non-Gaussian observation noise state space modeling. Here, we show that Markov switching models can be expressed in the general state space model form. A state estimation of the general state space model can be achieved by non-Gaussian filtering and smoothing algorithms (Kitagawa, 1987).

1.1. General State Space Model

Consider the system described by a general state space model,

$$\text{system equation} \qquad Z_n \sim q(Z_n|Z_{n-1}), \tag{1}$$

$$\text{observation equation} \qquad R_n \sim r(R_n|Z_n), \tag{2}$$

where R_n is the observed time series and Z_n is the unknown state vector at discrete time n. q and r are conditional distribution of Z_n given Z_{n-1} and of R_n given Z_n, respectively. The initial state vector Z_0 is distributed according to the distribution $p(Z_0|R_0)$ (Kitagawa and Gersch, 1996).

1.2. Markov Switching Model

In our research, we focus on price and unit sales which are common variables in a wide variety of POS data. The time series data of price and unit sales are indicated by x_n and y_n ($n = 1, \ldots, N$), in this study, respectively.

Markov Switching Distribution Model: MSD Model We assume that at time n, an observed time series x_n is generated by one of the k distribution models, where the choice of the distribution model is specified by the state. That is, denote a set of states by $l = \{1, \ldots, k\}$ and a discrete state variable at time n by S_n. We assume that when $S_n = i$, the distribution of x_n is given as follows:

$$g\big(x_n|S_n = i, \psi^x_{n-1}\big) = g_i(x_n), \tag{3}$$

where $\psi^x_{n-1} = \{x_1, \ldots, x_{n-1}\}$. A superscript x stands for the observation of x, explicitly. Furthermore, we assume that switching of state S_n follows a Markov chain with transition matrix given by

$$P = (p_{ij}) = \begin{pmatrix} p_{11} & \cdots & p_{1k} \\ \vdots & \ddots & \vdots \\ p_{k1} & \cdots & p_{kk} \end{pmatrix}, \tag{4}$$

where $p_{ij} = \Pr(S_n = j|S_{n-1} = i)$ with a constraint $\sum_{i=1}^{k} p_{ij} = 1$. In this case, the MSD model can be represented by the general state space form since equation (4) is a system model and equation (3) is an observation model.

Markov Switching Regression Model: MSR Model We assume that at time n, the observed time series y_n is generated by one of the k regression models, where the specification of the regression model is given by a discrete state variables. That is, we assume that when $S_n = i$, the distribution of y_n is given as follows:

$$f\big(y_n|S_n = i, \psi^y_{n-1}, x_n\big) = f_i\big(y_n|\mu_i(x_n, S_n = i)\big), \tag{5}$$

where $\psi^y_{n-1} = \{y_1, \ldots, y_{n-1}\}$ and x_n is treated as an exogenous variable. Furthermore, we assume that $\mu_i(x_n, S_n = i)$ is a parameter vector for describing distribution f_i, and an expectation which is constructed according to the Generalized Linear Model (GLM) (McCullagh and Nelder, 1989). Since equations (4) and (5) can be interpreted as a system model and observation model, respectively, the MSR model can be represented by the general state space form.

Markov Switching Bivariate Model: MSB Model We assume that at time n, an observed bivariate time series, $Y_n = (x_n, y_n)$, is generated by one of the k bivariate models, where the choice of the bivariate model is specified by a discrete state variable. In addition, we assume that when $S_n = i$, a bivariate distribution of Y_n can be factorized as follows.

$$\begin{aligned} h\big(x_n, y_n | S_n = i, \psi^{xy}_{n-1}\big) &= f\big(y_n | S_n = i, \psi^{xy}_{n-1}, x_n\big) g\big(x_n | S_n = i, \psi^{xy}_{n-1}\big) \\ &= f_i\big(y_n | \mu_i(x_n, S_n = i)\big) g_i(x_n), \end{aligned} \tag{6}$$

where $\psi^{xy}_{n-1} = \{Y_1, \ldots, Y_{n-1}\}$. A combination of (6) with equation (4) allows us to represent the MSB model by the general state space model.

When bivariate time series data (x_n, y_n) is given, the three models mentioned above rely on only the distribution information of variable x_n, of variable y_n and of both variables, respectively. Correspondingly, R_n in equation (2) is $R_n = \{x_n\}$, $R_n = \{y_n\}$, and $R_n = \{x_n, y_n\}$, respectively.

In the three Markov switching models (MSD, MSR and MSB models), stationary probability is easily calculated since the switching mechanism is independent of the observed time series data. Denote the probability that a state is in i at time n by $q_{i,n}$. According to equation (4), we can obtain the following equation,

$$\begin{pmatrix} q_{1,n} \\ \vdots \\ q_{k,n} \end{pmatrix} = \begin{pmatrix} p_{11} & \cdots & p_{k1} \\ \vdots & \ddots & \vdots \\ p_{1k} & \cdots & p_{kk} \end{pmatrix} \begin{pmatrix} q_{1,n-1} \\ \vdots \\ q_{k,n-1} \end{pmatrix}, \tag{7}$$

where $q_{1,j} + \cdots + q_{k,j} = 1$. The stationary probability, $\Pr(S_j = i) = q_i$ is calculated from equation (7) given $q_i = q_{i,n} = q_{i,n-1}$.

Phenomenon in which latent probability structure changes can be easily modeled using three Markov switching models with an arbitrary distribution form of f and g. A general state space model form allows for a useful recursive formula for state estimation, non-Gaussian filtering and smoothing algorithms (Kitagawa, 1987). Here a problem of state estimation is to evaluate the conditional distribution $\Pr(S_n | \Psi_m)$ which is the probability of S_n given the observation $\Psi_m = \{R_1, \ldots, R_m\}$. Here the state prediction, filtering and smoothing problems respectively refer to the situations in which, $m < n$, $m = n$ and $m > n$. We give a brief summary of recursive formulas.

- **One step ahead prediction**

$$\begin{aligned} \Pr(S_n = i | \Psi_{n-1}) &= \sum_{j=1}^{k} \Pr(S_n = i, S_{n-1} = j | \Psi_{n-1}) \\ &= \sum_{j=1}^{k} \Pr(S_n = i | S_{n-1} = j) \Pr(S_{n-1} = j | \Psi_{n-1}) \\ &= \sum_{j=1}^{k} p_{ji} \Pr(S_{n-1} = j | \Psi_{n-1}), \end{aligned} \tag{8}$$

- **Filtering**

$$\begin{aligned}\Pr(S_n = i|\Psi_n) &= \Pr(S_n = i|\Psi_{n-1}, R_n)\\ &= \frac{p(R_n|S_n = i, \Psi_{n-1})\Pr(S_n = i|\Psi_{n-1})}{p(R_n|\Psi_{n-1})},\end{aligned} \tag{9}$$

where $p(R_n|\Psi_{n-1})$ is obtained through $p(R_n|\Psi_{n-1}) = \sum_{i=1}^{k} p(R_n|S_n = i, \Psi_{n-1}) \times \Pr(S_n = i|\Psi_{n-1})$.

- **Smoothing**

$$\begin{aligned}\Pr(S_n = i|\Psi_N) &= \sum_{j=1}^{k} \Pr(S_{n+1} = j|\Psi_N)\Pr(S_n = i|S_{n+1} = j, \Psi_N)\\ &= \sum_{j=1}^{k} \Pr(S_{n+1} = j|\Psi_N)\Pr(S_n = i|S_{n+1} = j, \Psi_n)\\ &= \sum_{j=1}^{k} \Pr(S_{n+1} = j|\Psi_N)\frac{\Pr(S_n = i|\Psi_n)\Pr(S_{n+1} = j|S_n = i, \Psi_n)}{\Pr(S_{n+1} = j|\Psi_n)}\\ &= \Pr(S_n = i|\Psi_n)\sum_{j=1}^{k} \frac{\Pr(S_{n+1} = j|\Psi_N)\Pr(S_{n+1} = j|S_n = i)}{\Pr(S_{n+1} = j|S_n = i)}.\end{aligned} \tag{10}$$

Here the second line in equation (10) is obtained by the fact such that $P(S_n = i|S_{n+1} = j, \Psi_N) = P(S_n = i|S_{n+1} = j, \Psi_n)$ for the general state space model. $\Pr(S_n|\Psi_m)$ is generally non-Gaussian; its estimation should be made by numerical methods. However, for a discrete state space model, the numerical method is not required, because a state variable requires a finite number of integers.

When the general state space model generally has an unknown parameter vector θ, our parameter estimation is carried out by maximizing the log-likelihood defined by

$$l(\theta|y) = \sum_{n=1}^{N} \log\bigl(p(R_n|\Psi_{n-1})\bigr). \tag{11}$$

Note that $p(R_n|\Psi_{n-1})$ appears as the denominator in (9). Therefore, the log-likelihood is obtained as the by-product of the non-Gaussian filter.

When there exist competing models with a different number of parameters, we can select the model by Akaike's Information Criterion (AIC) (Akaike, 1974; Kitagawa, 1987),

$$\mathrm{AIC} = -2\max\bigl(l(\hat{\theta})\bigr) + 2(\text{number of free parameters}), \tag{12}$$

or Bayesian Information Criterion (BIC) (Schwarz, 1978),

$$\mathrm{BIC} = -2\max\bigl(l(\hat{\theta})\bigr) + (\text{number of free parameters}) \cdot \log(N). \tag{13}$$

2. Data Analysis

In this section, we apply the three Markov switching models to an analysis of brand in an instant coffee category. We examine weekly, store level, POS data and attempt to estimate the execution/non-execution of display promotion in a store using only POS data. Furthermore, we use the supervised data to check the estimation result, where the supervised data is data obtained by an actual marketing survey of the execution/non-execution of display promotion in a store.

2.1. Data

We investigate stochastic switching models using a subset of the DEI[1] weekly POS data. Information on weekly units and weekly sales for each brand and on the weekly number of visitors to the store were available. The method presented was applied to about six and a half year's instant coffee data for the period of 24 April 1995 to 1 December 2001 from one store of a well-known Japanese supermarket chain in Tokyo. The total number N is 345. Unfortunately, the chain and brand name cannot be shown clearly in this paper because of the data use contract. Average sales of groceries and the number of visitors per day of the store are about seven million yen and three thousand, respectively. The store is located in a closed trading area, and there are no competing stores. There are 77 items in the instant coffee category. The accumulated share of the top 10 units sold in this category is 67%. The target brand is a national brand, which gets a top share in the store of 16% during the period. The size attribute of the brand is 250 g. Prices of the brand ranged from 535 yen ($\cong$ \$4.5) to 1498 yen ($\cong$ \$12.5). The brand was displayed in the store, on average, about every two weeks and was shown on the flyer about every seven weeks. When the brand was shown on the flyer, the display promotion of the brand was executed with an extremely high possibility (about 98%).

In our analysis, we analyze two variables, Purchase Incidence (PI) and normalized price. Purchase incidence, y_n is obtained by

$$y_n = \frac{\mathrm{UNIT}_n \times 1000}{\mathrm{CUSTOMER}_n}, \tag{14}$$

where UNIT_n and $\mathrm{CUSTOMER}_n$ denote unit sales and visitors to a store at time n, respectively. Here, equation (14) indicates the unit sales of 1000 visitors to a store at time n. The normalized price, x_n is given by

$$x_n = 0.001 + \frac{(\mathrm{PRICE}_n - \mathrm{PRICE}_{\min})}{(\mathrm{PRICE}_{\max} - \mathrm{PRICE}_{\min})}(0.999 - 0.001), \tag{15}$$

where PRICE_n indicates the price at time n and $\mathrm{PRICE}_{\min}$ and $\mathrm{PRICE}_{\max}$ denote the minimum and maximum prices during the data period, respectively. For the sake of convenience, the values of 0.001 and 0.999 are introduced to fit the beta distribution to the distribution of normalized price.

In this paragraph, we define states. Here, we introduce a dummy variable ϕ_n (Table 1) to indicate the execution/non-execution of display promotion in a store at time n (execution 1,

Table 1. Features of Models

Number of regime		Regime	Display promotion		Number of parameter		
					MSB	MSR	MSB
Type 1 (2-states)	$S_n = 1$	Regime 1	$\phi_n = 0$		6	8	12
	$S_n = 2$	Regime 2	$\phi_n = 1$				
Type 2 (3-states)	$S_n = 1$	Regime 1	$\phi_n = 0$		9	12	18
	$S_n = 2$	Regime 2	$\phi_{n-1} = 0$	$\phi_n = 1$			
	$S_n = 3$	Regime 3	$\phi_{n-1} = 1$	$\phi_n = 1$			
Type 3 (4-states)	$S_n = 1$	Regime 1	$\phi_{n-1} = 0$	$\phi_n = 0$	12	16	24
	$S_n = 2$	Regime 2	$\phi_{n-1} = 0$	$\phi_n = 1$			
	$S_n = 3$	Regime 3	$\phi_{n-1} = 1$	$\phi_n = 0$			
	$S_n = 4$	Regime 4	$\phi_{n-1} = 1$	$\phi_n = 1$			

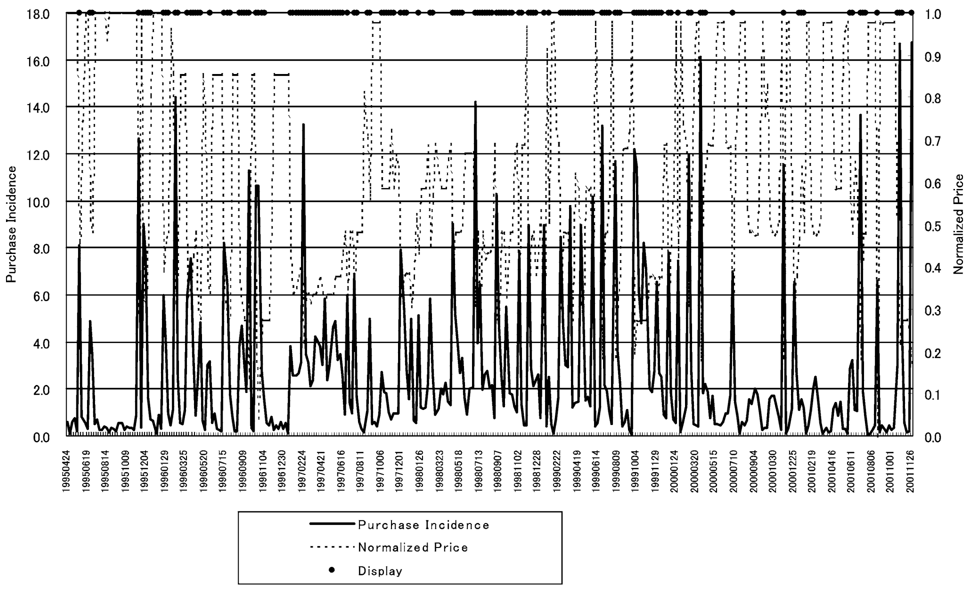

Figure 1. Time Series of Purchase Incidence and Normalized Price.

non-execution 0). A state can be defined by information on ϕ_n and/or ϕ_{n-1}. In the case of four states (= Type 3), the definition of state is as follows. Regime 1 represents the state where display promotion was not executed in the last week nor in this week ($\phi_n = \phi_{n-1} = 0$). The other three regimes for Type 3 are defined in Table 1.

For preliminary analysis, we first investigate POS data with supervised data. Figure 1 shows the time series of purchase incidence and normalized price. Figure 2 shows histograms of normalized price for every regime in the case of four states (= Type 3). The upper row in Figure 2 corresponds to the non-execution of display promotion in this week, and the lower rows correspond to the execution of display promotion in this week. The

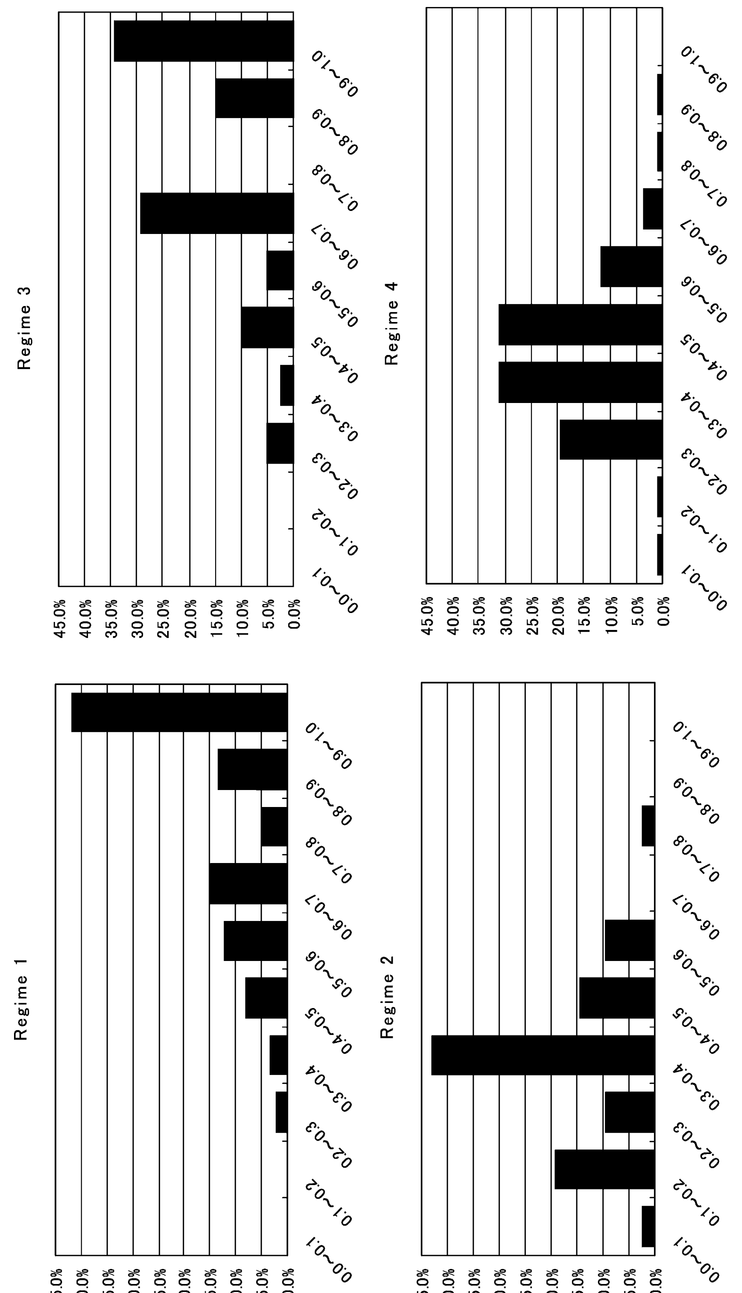

Figure 2. Histogram of Normalized Price (Every Regime for Type 3).

difference between upper and lower distributions is obvious in this figure. For example, the upper distributions are skewed to high price; in contrast, the lower ones are skewed to low price. Figure 3 shows a scatter plot of normalized price against purchase incidence for every regime of Type 3. The definition of regime is the same as in Figure 2. In upper distributions, the normalized price is mostly distributed between 0.5 and 1.0, and the variance of purchase incidence is small. On the contrary, in lower distributions, the normalized price is mostly distributed between 0.2 and 0.6, and the variance of purchase incidence is large. As this prior analysis showed, it turns out that there is a difference in a consumer's response to price, and the price strategy of a supplier side for every regime.

Table 2 demonstrates a switching matrix to specify whether the execution/non-execution of display promotion switches or not between last and this week. When the display promotion was not executed in the last week, the display promotion tends to be in the non-execution state (78%). On the other hand, when the display promotion was conducted in the last week, the display promotion is likely to be in the execution state (73%). This fact shows that the execution/non-execution of display promotion follows a dynamic mechanism, as shown above. It is therefore useful to take these dynamics into account in the statistical modeling.

2.2. *Assumptions and Concept*

Based on visual inspections in the previous subsection, we assume the following for modeling. Three assumptions are made for that purpose. First, we regard past unobserved display promotion as states, previously defined. Each state variable evolves according to an unknown transition probability matrix between states. Second, the purchase incidence is influenced by normalized price and the relation between purchase incidence and normalized price is different in every regime. Third, the distribution of normalized price is different in every regime.

Figure 4 illustrates the concept of our proposed model for Type 3. Here, we consider a hidden 4-states stochastic switching model. S_n is a latent variable which corresponds to a regime previously defined. When S_n is equal to 1, then it means regime 1, and so forth. If S_{n-1} exists in regime 1, then S_n will exist in regime 1 with probability p or moves to regime 2 with probability $1 - p$. But if S_{n-1} exists in regime 1, then S_n cannot exist in regime 3 and regime 4. As shown in this figure, other regimes are specified similarly. This system model shows the latent evolving mechanism of display promotion.

Moreover, we assume that at any time n, the observed time series is generated by one of three types of observation models for each regime: Markov Switching Distribution model, Markov Switching Regression model, or Markov Switching Bivariate model. These observation models indicate the observed characteristic for each regime.

2.3. *Model Specification*

In this section, we apply the three Markov switching models (MSD, MSR and MSB models) to POS data for the three types of regime definitions, listed in Table 1 (Types 1, 2

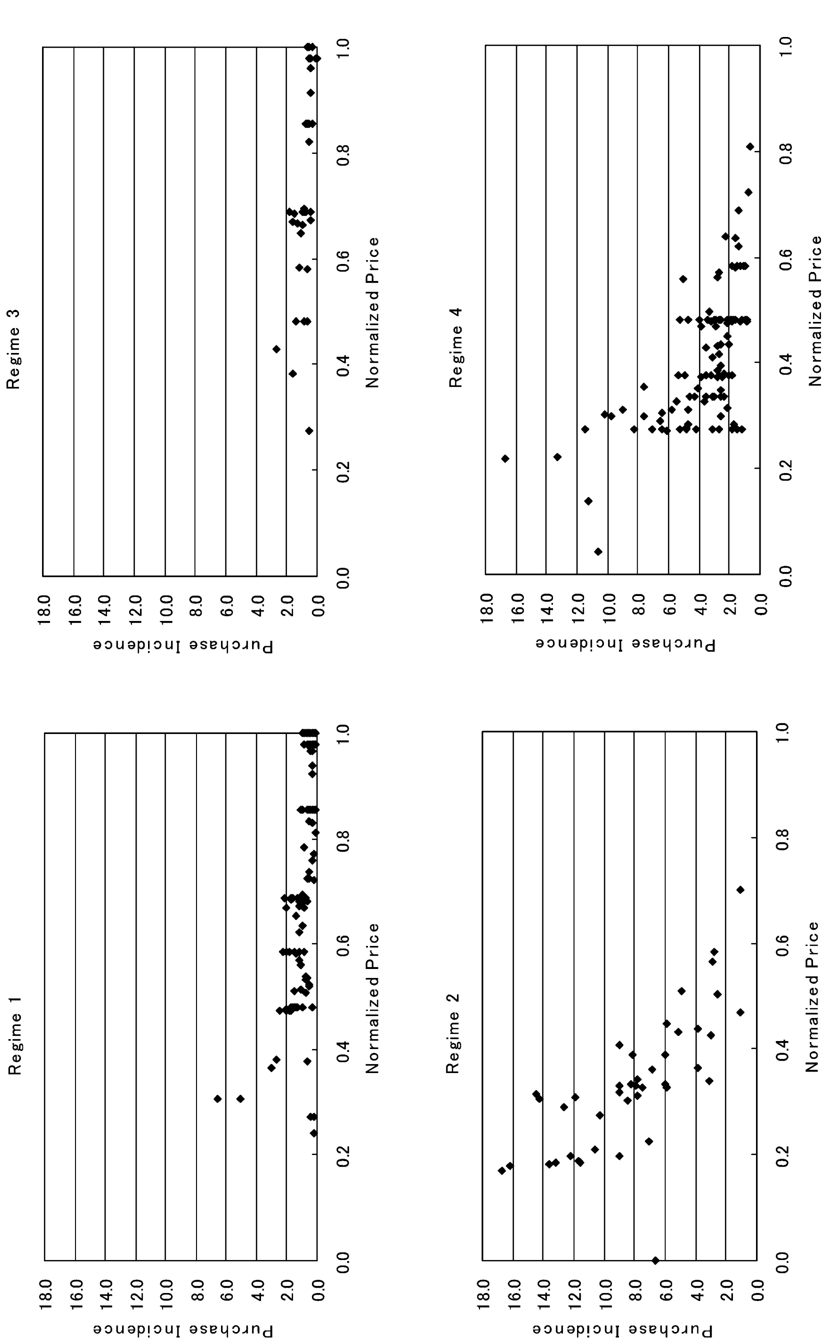

Figure 3. Scatter Plot (Every Regime for Type 3).

Table 2. Dynamic Features of Display

This week	Last week	
	Non-execution	Execution
Non-execution	0.7801	0.2662
Execution	0.2199	0.7338

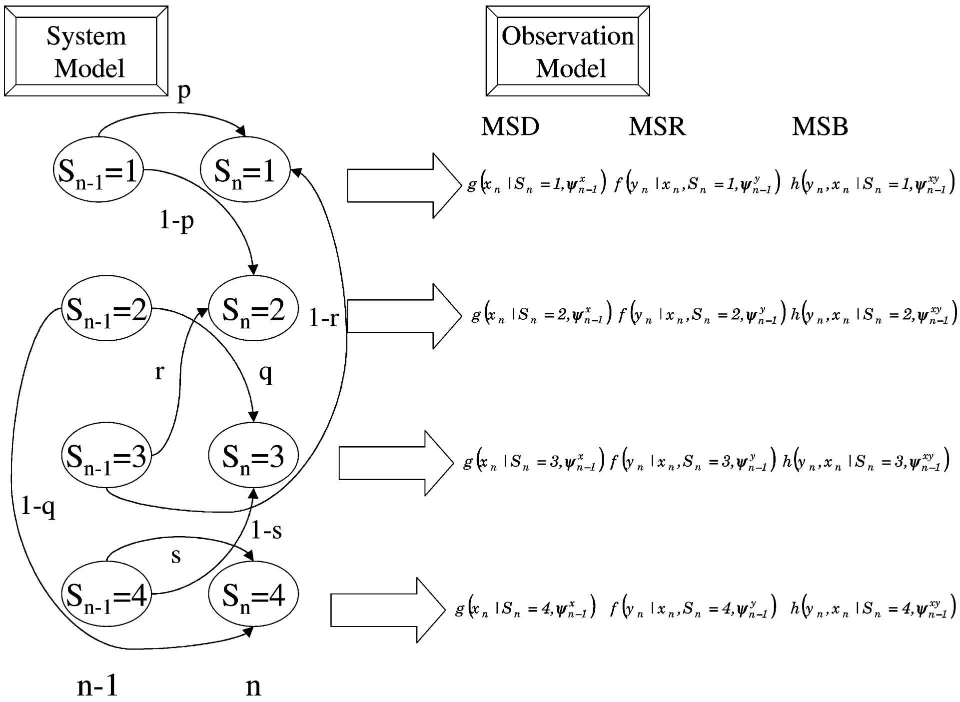

Figure 4. Conceptual Model.

and 3). Then, nine models in this study are applied to the same data set. However, the supervised data is not used for estimating these models. We explain the observation and system models for every type of regime definition.

Observation Model: MSD Model x_n is assumed to follow a beta distribution:

$$g\left(x_n | S_n = i, \psi^x_{n-1}\right) = g_i(x_n) = \frac{x_n^{\gamma_i - 1}(1 - x_n)^{\delta_i - 1}}{B(\gamma_i, \delta_i)}, \quad i = 1, \ldots, k, \tag{16}$$

where $B(\gamma_i, \delta_i)$ shows a beta function. We set $b_{1,i} = \log(\gamma_i)$, $b_{2,i} = \log(\delta_i)$ for parametrization. The distribution of normalized price, $g_i(x_n)$ depends on each regime.

Observation Model: MSR Model For the MSR model, we adopt the following model:

$$\begin{aligned} f\left(y_n|S_n = i, \psi^y_{n-1}, x_n\right) &= f_i(y_n|\mu_i) \\ &= \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left\{-\frac{1}{2\sigma_i^2}(y_n - \mu_i)^2\right\}, \quad i = 1, \dots, k, \end{aligned} \tag{17}$$

where μ_i is assumed in the linear combination as $\mu_i = \beta_{i0} + \beta_{i1}x_n$. We set $a_i = \log(\sigma_i^2)$. It should be noted that y_n is an observation and x_n is treated as an exogenous variable.

Observation Model: MSB Model For the MSB model, we use the following model:

$$\begin{aligned} &h\left(x_n, y_n|S_n = i, \psi^{xy}_{n-1}\right) \\ &\quad = f\left(y_n|S_n = i, \psi^{xy}_{n-1}, x_n\right) g\left(x_n|S_n = i, \psi^{xy}_{n-1}\right) = f_i(y_n|\mu_i) g_i(x_n) \\ &\quad = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left\{-\frac{1}{2\sigma_i^2}(y_n - \mu_i)^2\right\} \frac{x_n^{\gamma_i - 1}(1 - x_n)^{\delta_i - 1}}{B(\gamma_i, \delta_i)}, \quad i = 1, \dots, k. \end{aligned} \tag{18}$$

A set of x_n and y_n is treated as an observation. The MSB model is a model that is obtained by multiplication of the MSD model by the MSR model.

From a marketing perspective, the MSD model can be interpreted to represent the distribution of the price strategy of a supplier. On the other hand, the MSR model aims at describing the distribution of consumer's response. Finally, the MSB model deals with both the distribution of supplier's price strategy and consumer's response.

A system model for each type of regime definition is given as follows:

- **System Model: Type 1 = 2-States**

$$P = \begin{pmatrix} p & 1 - q \\ 1 - p & q \end{pmatrix}. \tag{19}$$

- **System Model: Type 2 = 3-States**

$$P = \begin{pmatrix} p & 1 - q & 1 - r \\ 1 - p & 0 & 0 \\ 0 & q & r \end{pmatrix}. \tag{20}$$

- **System Model: Type 3 = 4-States**

$$P = \begin{pmatrix} p & 0 & 1 - r & 0 \\ 1 - p & 0 & r & 0 \\ 0 & q & 0 & 1 - s \\ 0 & 1 - q & 0 & s \end{pmatrix}. \tag{21}$$

In addition, transition probabilities are parameterized with c_1, c_2, c_3, and c_4 as follows:

$$p = \frac{\exp(c_1)}{1 + \exp(c_1)},$$

Table 3. Estimated Parameter (2-States Models)

Parameters	MSD		MSR		MSB	
	Parameter	S.E.	Parameter	S.E.	Parameter	S.E.
a_1	–		−1.68721	(0.1359)[a]	−2.52719	(0.149)
a_2	–		2.28862	(0.1523)	2.10848	(0.1202)
$b_{1,1}$	0.35558	(0.1523)	–		0.83904	(0.1691)
$b_{2,1}$	−1.00860	(0.1044)	–		−0.73378	(0.1899)
$b_{1,2}$	1.68678	(0.1219)	–		1.43575	(0.745)
$b_{2,2}$	1.91547	(0.1365)	–		1.87633	(1.7878)
c_1	−0.19372	(0.2837)	−0.27769	(0.1787)	−0.30138	(0.1364)
c_2	−0.14297	(0.2405)	−0.62212	(0.2215)	−0.28803	(0.0904)
$\beta_{1,0}$	–		3.23569	(0.1357)	2.38257	(0.1095)
$\beta_{1,1}$	–		−3.01601	(0.1675)	−2.08213	(0.1158)
$\beta_{2,0}$	–		11.52666	(1.2462)	11.51955	(0.1912)
$\beta_{2,1}$	–		−15.49725	(3.2826)	−17.80411	(0.2072)

[a]Standard Error in parentheses.

$$q = \frac{\exp(c_2)}{1 + \exp(c_2)},$$
$$r = \frac{\exp(c_3)}{1 + \exp(c_3)},$$
$$s = \frac{\exp(c_4)}{1 + \exp(c_4)}.$$

In the 4-states MSB model, for example, unknown parameters are $\theta = \{\beta_{1,0}, \beta_{1,1}, a_1, \ldots, \beta_{4,0}, \beta_{4,1}, a_4, b_{1,1}, b_{2,1}, \ldots, b_{1,4}, b_{2,4}, c_1, c_2, c_3, c_4\}$.

2.4. *Model Comparison*

Estimated parameter values for nine models in Table 1 are listed in Tables 3, 4 and 5. Tables 3, 4 and 5 correspond to the results of 2-states, 3-states and 4-states models, respectively.

Table 6 shows the log-likelihood, AIC and BIC for nine models listed in Table 1. AIC prefers the 4-states model irrespective of the observation model. However, the result obtained by BIC differs from that by AIC for the MSD and MSB models. In the MSR model, AIC and BIC give the same results. Whereas in the MSD model, BIC prefers the 2-states model, it prefers the 3-state model in the MSB model. Since likelihood functions for MSD, MSR and MSB models are defined on normalized price, purchase incidence and a set of both, respectively, neither AIC nor BIC can be applied to compare the MSD, MSR and MSB models. Therefore, we consider the model's misclassification rate of nine models for choosing the best model among the nine models.

Although the 3- and 4-states model give the state probability against each of the three and four regimes, respectively, we focus on discrimination between the execution and non-execution of display promotion in this week. Therefore, we employ the smoothed marginal probability of 3- and 4-states probability given as follows:

Table 4. Estimated Parameter (3-States Models)

Parameters	MSD		MSR		MSB	
	Parameter	S.E.	Parameter	S.E.	Parameter	S.E.
a_1	–		−1.71108	(0.1263)[a]	−2.17151	(0.1124)
a_2	–		1.98255	(0.2424)	1.98295	(0.2203)
a_3	–		1.83153	(0.1817)	1.63884	(0.1283)
$b_{1,1}$	0.43306	(0.1661)	–		0.79387	(0.1547)
$b_{2,1}$	−1.03100	(0.1004)	–		−0.69129	(0.1795)
$b_{1,2}$	0.78352	(0.4426)	–		0.72705	(1.0972)
$b_{2,2}$	0.83439	(0.612)	–		1.31214	(2.7733)
$b_{1,3}$	1.72553	(0.056)	–		2.10561	(0.781)
$b_{2,3}$	1.95467	(0.0551)	–		2.58054	(2.0041)
c_1	1.45398	(0.2627)	−0.25168	(0.1891)	−0.27922	(0.124)
c_2	12.06047	(0.0958)	−0.52433	(0.4557)	−0.44599	(0.0891)
c_3	1.76364	(0.2366)	−0.44857	(0.3585)	−0.27334	(0.1365)
$\beta_{1,0}$	–		3.27439	(0.1192)	2.76762	(0.1587)
$\beta_{1,1}$	–		−3.05738	(0.1502)	−2.48795	(0.1437)
$\beta_{2,0}$	–		14.38064	(0.9759)	15.45288	(0.1496)
$\beta_{2,1}$	–		−17.56026	(2.5852)	−21.21243	(0.1995)
$\beta_{3,0}$	–		7.72835	(0.924)	9.40106	(0.3762)
$\beta_{3,1}$	–		−9.88791	(2.2203)	−15.24700	(0.2228)

[a]Standard Error in parentheses.

- **Smoothed Marginal Probability: 3-States**

$$\Pr(\phi_n = 0|\psi_N) = \Pr(S_n = 1|\psi_N), \tag{22}$$

$$\Pr(\phi_n = 1|\psi_N) = \Pr(S_n = 2|\psi_N) + \Pr(S_n = 3|\psi_N). \tag{23}$$

- **Smoothed Marginal Probability: 4-States**

$$\Pr(\phi_n = 0|\psi_N) = \Pr(S_n = 1|\psi_N) + \Pr(S_n = 3|\psi_N), \tag{24}$$

$$\Pr(\phi_n = 1|\psi_N) = \Pr(S_n = 2|\psi_N) + \Pr(S_n = 4|\psi_N). \tag{25}$$

Here, equations (22) (or (24)) and (23) (or (25)) indicate the smoothed marginal probability that display promotion does not execute, and does execute in this week, respectively. Figure 5 shows the smoothed marginal probability for each observation model of 4-states. The MSR model generates many switches between regimes, but the MSD model tends to keep on the same regime. The MSB model has a middle-aspect between the MSD and MSR model. 2-and 3-states models also have the same tendency as the 4-states case.

According to this smoothed marginal probability, we discriminate between the execution and non-execution of display promotion. The discrimination formula is defined by

$$D_n = \begin{cases} 1, & \text{if } P(\phi_n = 1|\psi_N) \geqslant 0.5 \text{ (execution)}, \\ 0, & \text{if } P(\phi_n = 1|\psi_N) < 0.5 \text{ (non-execution)}. \end{cases} \tag{26}$$

The rate of coincidence between the estimator, D_n and the observation of supervised data O_n (execution 1, non-execution 0) is evaluated by

Table 5. Estimated Parameter (4-States Models)

Parameters	MSB		MSR		MSB	
	Parameter	S.E.	Parameter	S.E.	Parameter	S.E.
a_1	–		−2.47575	(0.0432)[a]	−2.39169	(0.1361)
a_2	–		2.20030	(0.2157)	1.97702	(0.2152)
a_3	–		−1.79541	(0.3191)	−0.86141	(0.3214)
a_4	–		1.79641	(0.1925)	1.24302	(0.1489)
$b_{1,1}$	0.39123	(0.1795)	–		0.92184	(0.0596)
$b_{2,1}$	−1.12950	(0.1128)	–		−0.79233	(0.0678)
$b_{1,2}$	0.95104	0.3771)	–		0.74219	(0.4362)
$b_{2,2}$	1.17083	(0.4786)	–		1.30097	(0.3777)
$b_{1,3}$	1.50459	(0.4927)	–		0.73845	(0.4964)
$b_{2,3}$	0.02445	(0.2977)	–		−0.00342	(0.5946)
$b_{1,4}$	1.70109	(0.0867)	–		2.21203	(0.2626)
$b_{2,4}$	1.93580	(0.0969)	–		2.67877	(0.5232)
c_1	1.25619	(0.2677)	−0.37702	(0.1995)	−0.34454	(0.1588)
c_2	−15.51731	(1.6045)	−0.83723	(0.4125)	−0.84561	(0.0943)
c_3	−12.07878	(0.0506)	−1.47991	(0.4706)	−1.54099	(0.2005)
c_4	1.80848	(0.2335)	−0.57269	(0.4058)	−0.31497	(0.2396)
$\beta_{1,0}$	–		2.29514	(0.1494)	2.47824	(0.2041)
$\beta_{1,1}$	–		−1.99052	(0.178)	−2.18049	(0.2546)
$\beta_{2,0}$	–		13.38754	(0.6779)	15.98382	(0.1268)
$\beta_{2,1}$	–		−16.57720	(1.177)	−22.47007	(0.1346)
$\beta_{3,0}$	–		4.15266	(0.2262)	2.93693	(0.2092)
$\beta_{3,1}$	–		−4.02592	(0.3347)	−2.60365	(0.3397)
$\beta_{4,0}$	–		7.81625	(0.6093)	8.35476	(0.4344)
$\beta_{4,1}$	–		−10.09200	(1.4669)	−12.88299	(0.2375)

[a] Standard Error in parentheses.

Table 6. Log-likelihood, AIC and BIC

Number	MSD			MSR			MSB		
of state	$l(\theta)$	AIC	BIC	$l(\theta)$	AIC	BIC	$l(\theta)$	AIC	BIC
2	173.49	−334.97	−311.91	−572.46	1160.92	1191.67	−360.95	745.91	792.03
3	175.98	−333.95	−299.36	−560.31	1144.61	1190.73	−325.39	686.78	755.96
4	182.31	−340.62	−294.50	−545.09	1122.18	1183.67	−312.08	672.17	764.41

$$\mathrm{RC} = \frac{\sum_{n=1}^{N} I(O_n = D_n)}{N}, \tag{27}$$

where N is sample size and $I(\cdot)$ is defined by,

$$I(\cdot) = \begin{cases} 1, & \text{if } O_n = D_n, \\ 0, & \text{if } O_n \neq D_n. \end{cases} \tag{28}$$

Table 7 summarizes the misclassification rate and model selection results by AIC and BIC for all models. In MSD, MSR and MSB models, the minimum misclassification models are 2-states, 4-states and 3-states models, respectively. The model with the minimum misclassification rate among all models is the 3-states MSB model.

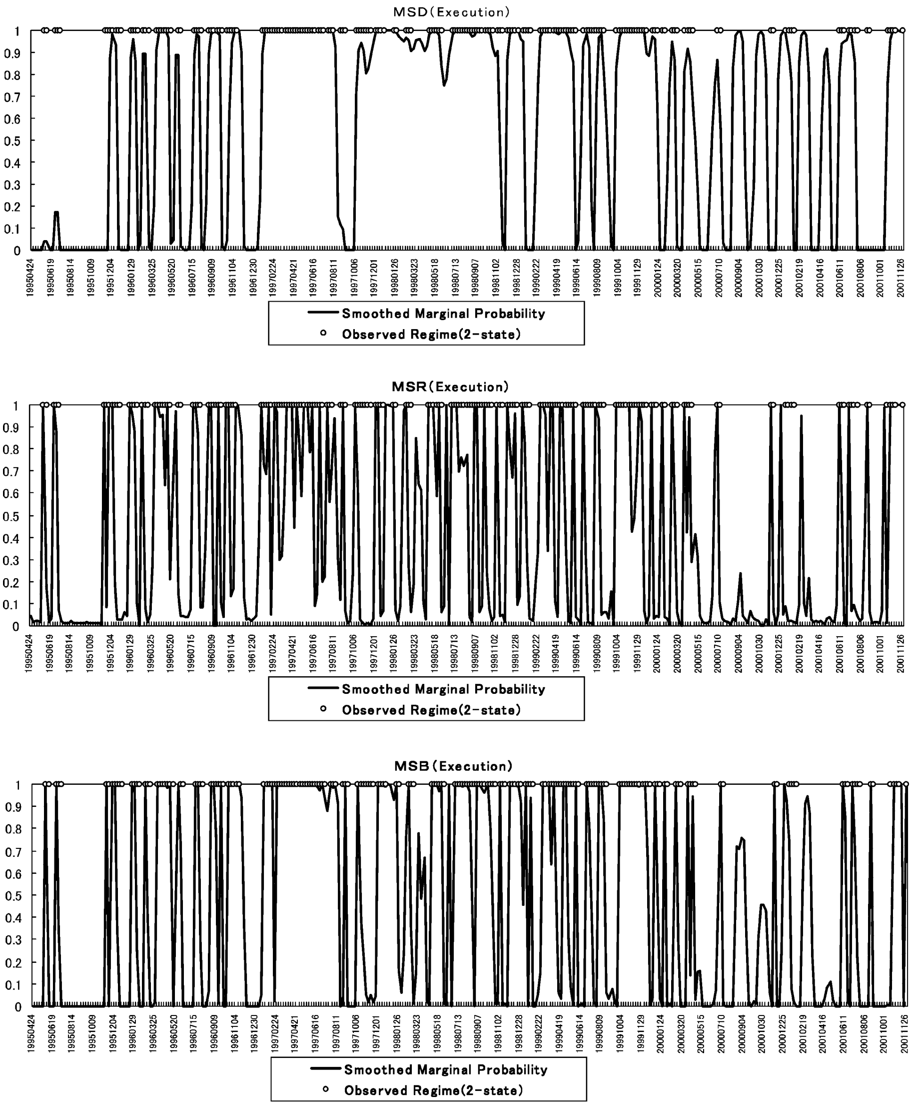

Figure 5. Marginal Smoothed Probability.

Here,we consider the relation between the misclassification rate and information criterion (i.e. AIC and BIC). As shown in Table 7, the minimum AIC model for MSD and MSB models does not correspond to the minimum misclassification rate model. The minimum BIC model for MSD, MSR and MSB models corresponds to the minimum misclassification rate model. Therefore, it is suggested that BIC is appropriate for discrimination using Markov switching models in this analysis. Thus, we focus on the minimum BIC models in subsequent discussion.

Table 7. Misclassification Rate (All Models)

Model	Variable	Non-execution	Execution	Total
2-States	Misclassification rate	0.35	0.10	0.24
MSD$^{(B)}$[b]	Number of misclassification sample	67	16	83
3-States	Misclassification rate	0.38	0.07	0.24
MSD	Number of misclassification sample	72	11	83
4-States	Misclassification rate	0.38	0.09	0.25
MSD$^{(A)}$[a]	Number of misclassification sample	73	14	87
2-States	Misclassification rate	0.06	0.42	0.22
MSR	Number of misclassification sample	12	65	77
3-States	Misclassification rate	0.08	0.36	0.21
MSR	Number of misclassification sample	15	56	71
4-States	Misclassification rate	0.11	0.31	0.20
MSR$^{(A),(B)}$	Number of misclassification sample	21	48	69
2-States	Misclassification rate	0.17	0.11	0.14
MSB	Number of misclassification sample	33	17	50
3-States	Misclassification rate	0.14	0.11	0.13
MSB$^{(B)}$	Number of misclassification sample	27	17	44
4-States	Misclassification rate	0.16	0.21	0.18
MSB$^{(A)}$	Number of misclassification sample	30	32	62

[a](A) indicates the minimum AIC model.
[b](B) indicates the minimum BIC model.

Figure 6 shows a scatter plot after the discrimination of two regimes for three minimum BIC models. The upper, middle and lower panels correspond to 2-states MSD, 4-states MSR and 3-states MSB models, respectively. In the 2-states MSD model, the non-execution samples, the normalized price of which is between 0.5 and 0.7, tend to be misclassified as the execution samples. The 4-states MSR model can identify these samples correctly, but the execution samples with purchase incidence between 2 and 4 are likely to be misclassified as the non-execution samples. The 3-states MSB model shows higher performance than 2-states MSD and 4-states MSR models in terms of discrimination.

3. Implication and Discussion

As we stated in previous sections, the execution/non-execution of display promotion was discriminated by Markov switching models without the supervised data. First, we showed that the MSB model is better than MSD and MSR models for all three types (2-, 3- and 4-states models) based on the results of the minimum misclassification rate. Next, we demonstrated that the 3-states MSB model among the three types of MSB models is the best in terms of the minimum misclassification rate (0.13). The misclassification rate, however, cannot be used for model selection, because the supervised data is not usually available. Therefore, another model selection procedure is required for that purpose. We showed that BIC can be used instead of the minimum misclassification rate.

We provide here more detailed analysis of the application result. Since the 3-states MSB model allows us to discriminate a display promotion correctly using only POS data, the effect of display promotion can be measured in the same manner as in prior literature, such as Blattberg et al. (1981), Neslin (2002). Once the effects of display promotion are obtained, suppliers can perform effective display promotion and price strategy at individual stores.

The fact that the 3-states MSB model was selected, implies that three typical patterns (regimes) exist for describing the relation between purchase incidence and normalized price. Similarly, the distribution of normalized price depends on each regimes. From the marketing perspective, it is suggested that both consumer's response to price (i.e. price sensitivity) and the supplier's price strategy differ among the three regimes. The price sensitivity in regime 1 is the lowest in the three regimes ($\beta_{1,1} = -2.48795$). On the other hand, the price sensitivity in regime 2 is the highest in the three regimes ($\beta_{2,1} = -21.21243$). The price sensitivity in regime 3 is in the middle of regime 1 and regime 2 ($\beta_{3,1} = -15.24700$). Since $\beta_{1,1,\text{w.s.}}$, $\beta_{2,1,\text{w.s.}}$, and $\beta_{3,1,\text{w.s.}}$[2] are -2.22981, -22.11850 and -13.77000, the price sensitivity for each regime can be accurately identified by our model-based approach. Therefore, accurate demand forecasting can be realized for individual stores using only POS data. There is a difference in the supplier's price strategy among the three regimes. The estimated average of normalized price in regime 1, regime 2 and regime 3 is 0.8154, 0.3578 and 0.38345, respectively. True values of these averages for every regime, which are calculated with the supervised data, are 0.7751, 0.3333 and 0.4121, respectively. A good agreement between the estimated and actual values indicates that the 3-states MSB model was able to detect the supplier's price strategy adequately.

To show the merit of our model-based approach, we compare the Radial Basis Function (RBF) network's[3] discrimination accuracy with that of the 3-states MSB model. Although details of the RBF networks are omitted in this paper, it is one of the most powerful nonlinear discriminant methods (Yee and Haykin, 2001). For this analysis, we divided our data into two parts. The RBF was learned by first half data (supervised data), and was evaluated by the misclassification rate of second half data. On the other hand, the 3-states MSB model was fitted to the second half data and the misclassification rate was calculated. Then, we compared the RBF's misclassification rate with that of the 3-states MSB model.

The misclassification rates of the RBF network and the 3-states MSB model are shown in Table 8. The 3-states MSB model's misclassification rate is lower than that of the RBF network. This result is very interesting, because while the RBF network needs the supervised data, our model-based approach does not. This empirical result clearly shows a benefit of incorporating the dynamic mechanism in a model-based approach.

Table 8. Misclassification Rate of RBF and 3-States MSB Model

Model	Variable	Non-execution	Execution	Total
RBF	Misclassification rate	0.15	0.19	0.17
	Number of misclassification sample	0.15	14	29
3-states	Misclassification rate	0.07	0.23	0.14
MSB	Number of misclassification sample	7	17	24

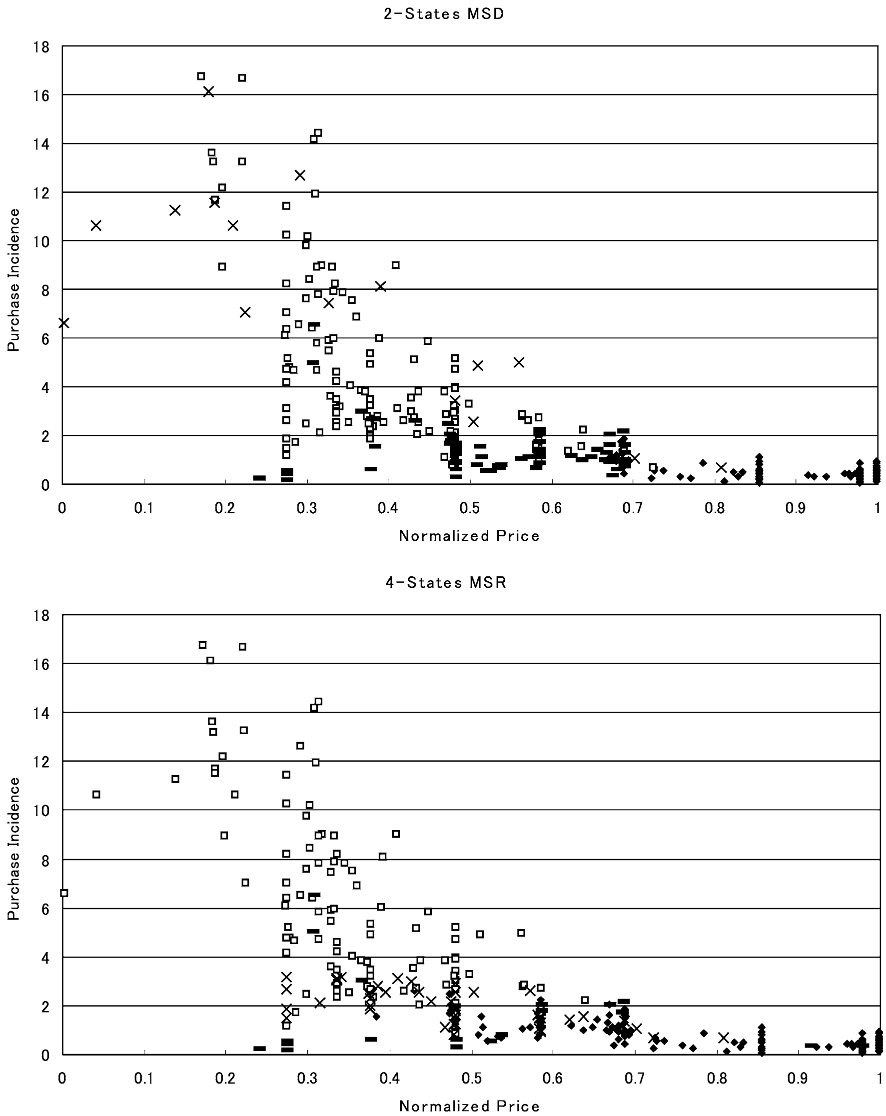

Figure 6. Scatter Plot of Discrimination Result.

Here, we mention three possibilities to reduce the misclassification rate. The first is associated with the sales units and price strategy of competing brands. From a marketing perspective, it is usually thought that two or more brands are competing. In our proposed model, however, since that information is not included, some samples may be misclassified according to influence due to competing brands. One way to deal with this problem is to develop a stochastic switching multivariate model that deals with a competing brand's

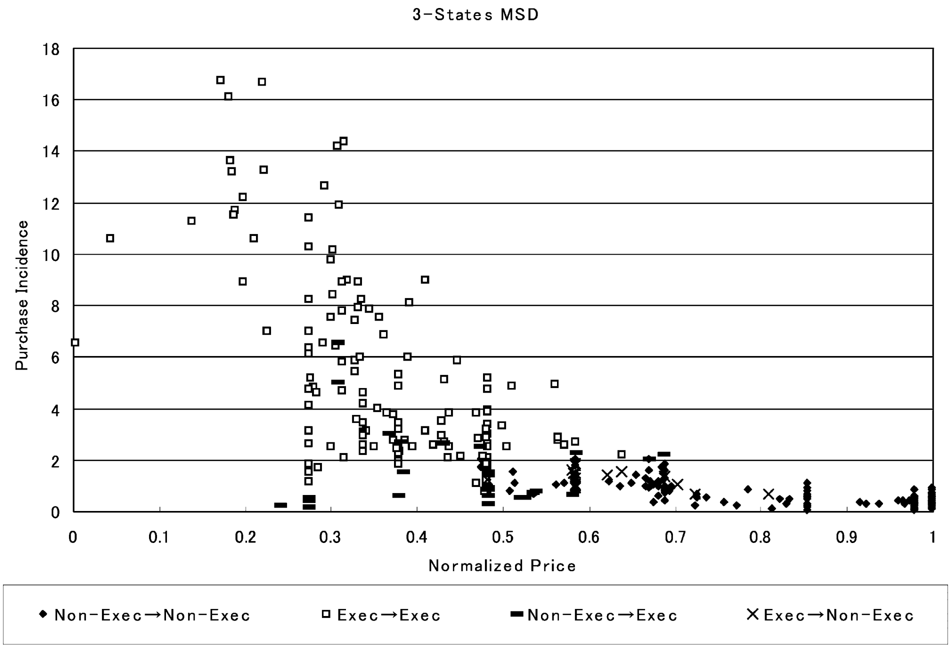

Figure 6. (Continued).

POS data simultaneously. The second way is to extend our linear observation model to a non-linear model. When these approaches are employed, many parameters must be optimized, which requires a reliable and robust method to estimate many parameters stably.

The third way is related to the structure of our used POS data. In our research, we apply Markov switching models to six and a half year's instant coffee data. The data structure itself may be changing even within this data period. When there is a structural change of the data, our proposed models cannot respond to accommodate such change. This issue can be dealt with by the framework of the time varying parameter model (Kitagawa and Gersch, 1985; West and Harrison, 1997; Kim and Nelson, 1999). Our proposed model can be extended in the same manner as these prior literatures assuming a gradual change in parameters.

We want to discuss the applicability of the Markov switching model proposed in this paper for various marketing issues. First, for example, we consider *category management*, which is one of the main issues in marketing. The important thing in this issue is to recognize whether the trend of a category is in expansion or recession terms. If the trend is treated as latent variables which switch between the expansion and recession states, this phenomenon can be formulized in the modeling framework of our proposed model.

Our model-based approach is also applicable to consumer segmentation. Usually, in the consumer segmentation carried out in marketing, each consumer belongs only to one segment throughout all time. However, it may be natural to suppose that a consumer switches between two or more segments in response to changes in a supplier's marketing activity or a consumer's own taste at some time. If a consumer segment as a stochastic time-dependent latent variable switches between latent segments, this dynamic segmentation can be realized in the framework of our proposed model.

4. Concluding Remarks

This paper considered the problem of identifying unobserved display promotions using only POS data. For that purpose, we first investigated POS data with supervised data, which has information on the execution/non-execution of in-store display promotions, and attempted to identify the dependency of an execution/non-execution of a display promotion on two quantities in POS data. Integration of this analysis with prior information could be realized by stochastic switching models with unknown state variables. The state estimation was carried out with non-Gaussian filtering and smoothing algorithms. We applied the proposed models to POS data without the supervised data, and obtained a highly accurate estimator of unobserved display promotion. The performance of our proposed model was tested by comparing it with RBF (supervised learning) and MSB (unsupervised learning). Our approach is attractive for the marketing field, because the unobserved display promotion was correctly discriminated without the supervised data. This feature can be realized by incorporating prior information that characterizes the dynamic behavior of unknown state variables.

The Markov switching model usually assumes an explicit dependency of a state on the observation value. For an example, previous works (cf. Blattberg and Wisniewski (1987), Wittink et al. (1987)) often assume an explicit analytical form or algorithmic rule for describing the relation between purchase incidence and parameters of the execution/non-execution of display promotion. Meanwhile, our model has a simple structure where the switches arise independently of the observation. However, our model-based approach can extract information on the relation between the state and observation, as a result of applications. Namely, our approach is capable of discriminating latent states effectively without assuming the explicit dependency.

In conclusions, the analysis of POS data with a time series model is effective for extracting marketing information, which indicates that model-based time series analysis is an important marketing tool. It should be noted that since the state variable takes a discrete value, the computational burden could be reduced much more than in a continuous case. These points are useful features of our model-based approach.

Acknowledgments

Work at the Institute of Statistical Mathematics was carried out in part under a Grant-in-Aid for Science Research (A) (14208025).

Notes

1. The Distribution Economics Institute was established in 1963 as a voluntary group. In 1966, under the guidance of the Ministry of International Trade and Industry (now Ministry of Economy, Trade and Industry), DEI was transformed into a nonprofit research organization, http://www.dei.or.jp/
2. These values were obtained through the linear regression model using the POS data with supervised data for each of the three regimes, w.s. stands for an analysis result with the supervised data.
3. The supervised learning.

References

Akaike, H. (1974). "A New Look at the Statistical Model Identification," *IEEE Trans. Autom. Contr.*, AC-19, 716–723.

Blattberg, Robert C. and Kenneth J. Wisniewski. (1987). "How Retail Price Promotions Work: Empirical Results," Marketing working paper No. 42, University of Chicago.

Blattberg, Robert C., Gary Eppen, and Joshua Liebermann. (1981). "A Theoretical and Empirical Evaluation of Price Deals in Consumer Non-Durables," *Journal of Marketing*, 45, 116–129.

Bucklin, Randolph E. and James M. Lattin. (1991). "A Two-State Model of Purchase Incidence and Brand Choice," *Marketing Science*, 10(1), 24–39.

Goldfeld, S. M. and R. E. Quandt. (1973). "A Markov Model for Switching Regression," *Journal of Econometrics*, 1, 3–16.

Gordon, K. and A. F. M. Smith. (1988). "Modeling and Monitoring Discontinuous Change in Time Series." In J. C. Spall (ed.), *Bayesian Analysis of Time Series and Dynamic Linear Model*. New York: Marcel Dekker, 359–392.

Gordon, K. and A. F. M. Smith. (1990). "Modeling and Monitoring Biomedical Time Series," *Journal of the American Statistical Association*, 85, 328–337.

Guadagni, Peter M. and John D. C. Little. (1983). "A Logit Model of Brand Choice Calibrated on Scanner Data," *Marketing Science*, 2, 203–238.

Gupta, Sunil. (1988). "Impact of Sales Promotions on When, What, and How Much to Buy," *Journal of Marketing Research*, 25.

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

Harrison, P. J. and C. F. Stevens. (1976). "Bayesian Forecasting (with Discussion)," *Journal of Royal Statistical Society Ser. B*, 38, 205–247.

Higuchi, T. and G. Kitagawa. (2000). "Knowledge Discovery and Self-Organizing State Space Model," *IEICE Transactions on Information and Systems*, E83-D(1), 36–43.

Kim, Chang-Jin and Charles R. Nelson. (1999). *State Space Models with Regime Switching*. Cambridge, MA: The MIT Press.

Kitagawa, G. (1987). "Non-Gaussian State-Space Modeling of Nonstationary Time Series," *Journal of the American Statistical Association*, 82, 1032–1063.

Kitagawa, G. and W. Gersch. (1985). "A Smoothness Priors Time Varying AR Coefficient Modeling of Nonstationary Time Series," *IEEE Trans. on Automatic Control*, AC-30, 303–314.

Kitagawa, G. and W. Gersch. (1996). *Smoothness Priors Analysis of Time Series*, Lecture Notes in Statistics, Vol. 116. New York: Springer.

McCullagh, P. and J. A. Nelder. (1989). *Generalized Linear models*, 2nd edition. London: Chapman and Hall.

Montgomery, Davit. (1971). "Consumer Characteristics Associated with Dealing: An Empirical Example," *Journal of Marketing Research*, 4, 118–120.

Neslin, Scott. (2002). *Sales Promotion*. Cambridge, MA: Marketing Science Institute.

Neslin, Scott, Caroline Henderson, and John Quelch. (1985). "Coupon Promotions and Acceleration of Product Purchase," *Marketing Science*, 4.

Pena, D. and I. Guttman. (1988). "Bayesian Approach to Robustifying the Kalman Filter." In J. C. Spall (ed.), *Bayesian Analysis of Time Series and Dynamic Linear Model*. New York: Marcel Dekker, 227–254.

Quandt, R. E. (1972). "A New Approach to Estimating Switching Regressions," *Journal of the American Statistical Association*, 67, 306–310.

Schwarz, G. (1978). "Estimating the Dimension of a Model," *Annals of Statistics*, 6, 461–464.

Tajima, Y. (ed.). (1989). *In-Store Merchandising* [in Japanese]. Tokyo: Bussiness-sha.

West, Mike and Jeff Harrison. (1997). *Bayesian Forecasting and Dynamic Models*, 2nd edition. New York: Springer.

Wittink, Dick R., Michael J. Addona, William J. Hawkes, and John C. Porter. (1987). "SCANPRO: A Model to Measure Short-Term Effects of Promotional Activities on Brand Sales, Based on Store-Level Scanner Data," working paper, Johnson Graduate School of Management, Cornell University, Ithaca, NY.

Yee, V. P. and S. Haykin. (2001). *Regularized Radial Basis Function Networks*. Wiley Inter-Science.