

マルコフ切換モデルによる観測されない非価格プロモーション実施の有無の統計的推測法

佐藤忠彦* 樋口知之** 北川源四郎**

* 筑波大学ビジネス科学研究科 ** 情報・システム研究機構 統計数理研究所

Statistical Inference using Stochastic Switching Models for the
Discrimination of Unobserved Non-Price Promotion

Tadahiko Sato* Tomoyuki Higuchi** Genshiro Kitagawa**

* Graduate School of Business Science, University of Tsukuba, Tokyo

** Research Organization of Information and Systems, The Institute of Statistical Mathematics

Abstract. It is known that an execution of price and/or a non-price promotion has a strong influence on the sales of a brand sold in a supermarket. Usually, we can easily obtain information on a price promotion from a POS data. On the other hand, unless investigator collects information on an execution of non-price promotion in every retail store, we can not obtain such information. In this paper, we consider a problem to identify whether or not non-price promotion is conducted. We treat a non-price promotion execution/non-execution 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. An objective of the analysis is to identify the model and to assign a probability model for each state at each time instant. Finally, we give a high precision estimator of a past non-price promotion based on the proposal model.

1 はじめに

マクロ経済データには経済変動の結果、発展期と後退期が交互に現れる。また、POS (point of sales) データと呼ばれるマーケティングの分野で活用される販売データでは、店内における価格プロモーションや、店内における大量陣列のような非価格プロモーションの実施の有無により、販売量の変動パターンが大きく変化する。このようないくつかの異なった変動パターンからなる時系列データは様々な分野で数多くみられる。

このようなデータの解析のために、確率構造が切り換わるモデルが考えられてきている。時系列解析では、[4, 5, 7, 8] は誤差分散の変動によって、[19] は観測誤差に混合分布を仮定することによって、確率構造が切り換わるモデルを表現している。[20] ではデザイン行列の切換によって、回帰モデルにおける確率構造の切換をモデル化している。[3] は独立な観測値に対する定常なマルコフ切換モデルによって経済が発展期か収縮期かを判定する問題を考えている。従属関係にある観測値を用いたマルコフ切換モデルとしては、[6] では経済の成長期と後退期々々を潜在的な状態として捉えて、それらのマルコフ切換によるモデル化を行っている。

一方、最近、工学の分野では音声認識などへの応用実績から、隠れマルコフモデル (Hidden Markov Model) と呼ばれるモデルが注目を浴びている。観測値が隠れマルコフ連鎖に支配される混合モデルにおいて、[10] では、EM アルゴリズムにより逐次的にパラメータ推定するアルゴリズムが提案されている。この隠れマルコフモデルと状態空間モデルは非常に関係が深い。多くの場合、これらのモデルを利用する目的は、観測値に基づく潜在的な状態の推定である。その他マルコフ切換モデルへ状態空間モデルを適用したものとしては、[11, 16,

23]などがある。これらの論文では、潜在的に仮定された状態間の遷移確率をシステムモデルとみなしてモデルを状態空間モデル表現し、[12]で提案された、非ガウス型フィルター及び平滑化で状態推定を行っている。

マーケティングにおいて、スーパーマーケットで販売されるブランドの売上には、価格プロモーションと呼ばれる値引きと、それ以外の非価格プロモーションの実施が強く影響することが知られている。通常、POSデータでは販売点数、販売金額、来店客数の情報は取得できる。そのため、通常売価に比較して値引きがどの程度されていたか、すなわち価格プロモーションの有無に関してはPOSデータから簡単に判断できる。一方、非価格プロモーションの有無については小売店舗毎に調査により収集しない限りはその情報が入手できない。しかし、非価格プロモーションの有無により大きく売上は変動するので、需要予測などを行う際には、非価格プロモーション実施時の構造を販売数量及び売価などのPOSデータ中の通常獲得できるデータからの的確に捉えることが予測精度の向上の視点からは非常に重要である。また、非価格プロモーションを実施していたかどうかを、販売点数、販売時点での売価及び来店客数のデータから判別できれば、施策の効果を踏まえてマーケティングに関する意思決定を行える。

非価格プロモーションのデータは、前述したように消費財メーカーはいうまでもなく、小売業においてさえ取得できていない（取得しようと思えしていない）現状にある。このような状況になる理由は、テレビ広告などに比べて値引きや非価格プロモーションの実施頻度は高く、しかもその内容が小売店舗毎に異なるために発生している。また、小売業が消費財メーカーにそれらを開示したくないといった理由もある。しかし、前述したように非価格プロモーションの実施は売上に強く影響する。また、小売業が非価格プロモーションを実施する場合、消費財メーカーは小売業へ協賛金などを支払わなければならないケースが多い。消費財メーカーが限られた販売促進費を有効に活用するためには、非価格プロモーションの効果を的確に測定することが必要である。そのため、数多くの店舗でこの非価格プロモーションデータを取得することが、消費財メーカーのひとつの大きなマーケティング課題であるといつても過言ではない。これら非価格プロモーションデータを店頭で調査により取得するためにかかる費用は、参考ではあるが年間400万円/1店舗程度かかる。消費財メーカーが、実際に数多くの店舗でこれだけの費用を支払ってデータを取り続けることは不可能であると考えられる。

そのために2節では、上記のような先行研究及び問題意識を踏まえた上で、教師データが存在しない下で時系列サンプルの各時点における状態を判別するためのモデルを考える。そこではマルコフ切換モデルを提案し、それを状態空間モデルを用いて統一的に扱う。3節では数値例として実際のPOSデータへモデルを適用しモデルの評価を行う。4節では、シミュレーションにより、モデルの動的特性・共変量の変動の影響に関して検証を行う。最後に5節ではまとめと今後の課題を示す。

2 マルコフ切換モデル

本節においては、通常の線形・ガウス型の状態空間モデルを非線形・非ガウスの状態及び観測過程まで取り扱えるように拡張された一般状態空間モデルを紹介し、マルコフ切換分布モデル、回帰モデル、2変量モデルが一般状態空間モデルを用いて表現できることを示す。状態ベクトルの状態推定は、非ガウス型フィルタ及び平滑化により履行される[12]。

(一般状態空間モデル)

時刻 n の観測された時系列ベクトルを R_n 、未知の状態ベクトルを Z_n とすると、以下に示す2つの式により一般状態空間モデルが定義される。

$$(2.1) \quad \text{状態方程式} \quad Z_n \sim q(Z_n | Z_{n-1}),$$

$$(2.2) \quad \text{観測方程式} \quad R_n \sim r(R_n | Z_n).$$

ただし, $q(Z_n|Z_{n-1})$ は Z_{n-1} を与えた場合の Z_n の条件付き分布, $r(R_n|Z_n)$ は Z_n を与えた場合の R_n の条件付き分布を示す.

ここで導入した一般状態空間モデルは線形・ガウス型の時系列や非線形時系列, 離散時系列などの様々な時系列データを統一的な枠組みで取り扱うことを可能にするモデルである [14, 15].

本研究では, POSデータの代表的な変数である売価と販売点数に注目する. 従ってここでは, 異なる種類の2つのスカラー時系列 $x_n, (n = 1, \dots, N), y_n, (n = 1, \dots, N)$ を考える.

[マルコフ切換分布モデル (Markov Switching Distribution Model) : MSDモデル]

時系列 x_n の分布は未知の状態 S_n に依存してモデルを切換えるものとする. すなわち状態 S_n がとりうる値の集合を $l = \{1, \dots, k\}$ とし, $S_n = i$ のとき時系列 x_n は下記の(2.3)式に従うものと仮定する.

$$(2.3) \quad g(x_n|S_n = i, \psi_{n-1}^x) = g_i(x_n).$$

ただし, $\psi_{n-1}^x = \{x_1, \dots, x_{n-1}\}$ とする. 上付の添字 x は, 変数 x の観測値であることを明示的に示す.

さらに, 状態 S_n の切換は下記(2.4)式の推移確率をもつマルコフ連鎖に従うものと仮定する.

$$(2.4) \quad Pr(S_n = i|S_{n-1} = j) = p_{ji}.$$

当然 $\sum_{i=1}^k p_{ij} = 1$ が成り立つ. このとき, (2.4)式をシステム方程式, (2.3)式を観測方程式とするとマルコフ切換分布モデルが(2.1), (2.2)式の一般状態空間モデルで表現できる.

[マルコフ切換回帰モデル (Markov Switching Regression Model) : MSRモデル]

時系列 y_n の分布は未知の状態 S_n に依存して切り換わり, その分布は時系列 x_n に影響されるものとする. 状態がとりうる値の集合を $l = \{1, \dots, k\}$ とすると, $S_n = i$ のとき時系列 y_n は下記(2.5)式に従うものと仮定する.

$$(2.5) \quad f(y_n|S_n = i, \psi_{n-1}^y, x_n) \equiv f(y_n|S_n = i, \psi_{n-1}^y, \mu_i) = f_i(y_n|\mu_i).$$

ただし, $\psi_{n-1}^y = \{y_1, \dots, y_{n-1}\}$ とする.

上記(2.5)式内の期待値 μ_i は一般化線形モデル (Generalized Linear Model, 以降GLM) の期待値の構成方法と同様に構成されるものとする [18]. このとき, (2.4)式をシステム方程式, (2.5)式を観測方程式とするとマルコフ切換回帰モデルの一般状態空間モデル表現が与えられる.

さらに, 上記で定義したMSDモデルとMSRモデルを融合することで, 下記に示すマルコフ切換2変量モデルが導出される.

[マルコフ切換2変量モデル (Markov Switching Bivariate Model) : MSBモデル]

2変量時系列 $Y_n = (x_n, y_n)$ の分布は未知の状態 S_n に依存して切り換わり, また時系列 y_n の分布は時系列 x_n に影響されるものとする. 状態がとりうる値の集合を $l = \{1, \dots, k\}$ とすると, $S_n = i$ のとき時系列 (x_n, y_n) は下記(2.6)式に従うものと仮定する.

$$(2.6) \quad h(x_n, y_n|S_n = i, \psi_{n-1}^{xy}) = f(y_n|S_n = i, \psi_{n-1}^{xy}, x_n)g(x_n|S_n = i, \psi_{n-1}^{xy}) = f_i(y_n|\mu_i)g_i(x_n).$$

ただし, $\psi_{n-1}^{xy} = \{Y_1, \dots, Y_{n-1}\}$ とする.

このとき, (2.4)式をシステム方程式, (2.6)式を観測方程式とするとマルコフ切換2変量モデルの一般状態空間モデル表現が与えられる.

2変量時系列 (x_n, y_n) が存在した場合, 上記に示した3つのモデルはそれぞれ, x_n の分布情報のみを, y_n の分布情報のみを, 両方の分布情報を活用しているモデルである.

上記で導入したマルコフ切換モデルは、切換が時系列の値とは独立に生じるため各状態の定常確率は簡単に求めることが出来る。時点 n での状態が $S_n = i, (i = 1, \dots, k)$ となる確率を $q_{i,n}$ と定義すると、

$$(2.7) \quad \begin{pmatrix} q_{1,n} \\ \vdots \\ q_{k,n} \end{pmatrix} = \begin{pmatrix} p_{11} & \cdots & p_{k1} \\ \vdots & \ddots & \vdots \\ p_{1k} & \cdots & p_{kk} \end{pmatrix} \begin{pmatrix} q_{1,n-1} \\ \vdots \\ q_{k,n-1} \end{pmatrix},$$

が成り立つ。ただし、 $q_{1,j} + \cdots + q_{k,j} = 1$ 。従って、状態の定常確率 $Pr(S_j = i) = q_i$ を求めるためには、 $q_i = q_{i,n} = q_{i,n-1}$ とおいて、上記 (2.7) 式を解けばよい。

上記で定義した 3 つのマルコフ切換モデルを用いることで、潜在的に確率構造が切り換わる現象を、様々な分布型に対して共変量を取り込んだ形で比較的容易にモデル化できる。これは、隠れマルコフモデル (Hidden Markov Model) の観測方程式を連続値版に拡張したモデルと考えると分かりやすいかもしれない [9]。

上記で定義されるマルコフ切換モデルはすべて一般状態空間モデルで統一的に記述できるので、以下に示すように非ガウス型フィルタのアルゴリズムで状態推定を行える [12]。なお、観測値 $\Psi_t = \{R_1, \dots, R_t\}$ に基づいて状態推定を考えた場合、 $t < n$ の場合は観測区間よりも先の状態を予測する問題で予測と呼ばれ、同様に $t = n$ の場合は現在の状態を推定する問題でフィルタと呼ばれる。また、 $t > n$ の場合は今までの観測値に基づいて過去の状態を推定する問題で平滑化と呼ばれる [13]。

(予測)

$$(2.8) \quad \begin{aligned} Pr(S_n = i | \Psi_{n-1}) &= \sum_{j=1}^k Pr(S_n = i, S_{n-1} = j | \Psi_{n-1}) \\ &= \sum_{j=1}^k Pr(S_n = i | S_{n-1} = j) Pr(S_{n-1} = j | \Psi_{n-1}) \\ &= \sum_{j=1}^k p_{ji} Pr(S_{n-1} = j | \Psi_{n-1}). \end{aligned}$$

(フィルタ)

$$(2.9) \quad \begin{aligned} Pr(S_n = i | \Psi_n) &= Pr(S_n = i | \Psi_{n-1}, R_n) \\ &= \frac{p(R_n | S_n = i, \Psi_{n-1}) Pr(S_n = i | \Psi_{n-1})}{p(R_n | \Psi_{n-1})}. \end{aligned}$$

ここで、 $p(R_n | S_n = i, \Psi_{n-1})$ は前述した 3 つのモデルでそれぞれ (2.3), (2.5), (2.6) 式で与えられ、 $p(S_n = i | \Psi_{n-1})$ は (2.8) 式で求められる。

上記 (2.9) 式内の分母 $p(R_n | \Psi_{n-1})$ は、 $n - 1$ 時点までの情報が与えられた場合の R_n の条件付き確率密度で、下記 (2.10) 式のように表現できる。

$$(2.10) \quad p(R_n | \Psi_{n-1}) = \sum_{i=1}^k p(R_n | S_n = i, \Psi_{n-1}) Pr(S_n = i | \Psi_{n-1}).$$

(平滑化)

次に平滑化の公式を与える。一般状態空間モデルに対して、 $P(S_n = i|S_{n+1} = j, \Psi_N) = P(S_n = i|S_{n+1} = j, \Psi_n)$ が成り立つことを利用すると、

$$\begin{aligned}
 Pr(S_n = i, S_{n+1} = j|\Psi_N) &= Pr(S_{n+1} = j|\Psi_N)Pr(S_n = i|S_{n+1} = j, \Psi_N) \\
 &= Pr(S_{n+1} = j|\Psi_N)Pr(S_n = i|S_{n+1} = j, \Psi_n) \\
 &= Pr(S_{n+1} = j|\Psi_N) \frac{Pr(S_n = i|\Psi_n)Pr(S_{n+1} = j|S_n = i, \Psi_n)}{Pr(S_{n+1} = j|\Psi_n)} \\
 (2.11) \quad &= Pr(S_n = i|\Psi_n) \frac{Pr(S_{n+1} = j|\Psi_N)Pr(S_{n+1} = j|S_n = i)}{Pr(S_{n+1} = j|\Psi_n)},
 \end{aligned}$$

が成り立つので、平滑化の分布は

$$(2.12) \quad Pr(S_n = i|\Psi_N) = Pr(S_n = i|\Psi_n) \sum_{j=1}^k \frac{Pr(S_{n+1} = j|\Psi_N)Pr(S_{n+1} = j|S_n = i)}{Pr(S_{n+1} = j|S_n = i)}.$$

で求められる。

一般の連続状態の状態空間モデルへ非ガウス型フィルタ及び平滑化の公式を適用するためには、状態空間を離散表現し、積分を和で近似する必要がある。しかし、離散状態の状態空間モデルの場合にはその必要性がなく、フィルタ及び平滑化後の確率も (2.8), (2.9), (2.12) のように簡単に計算できる。

上記で示したマルコフ切換モデルには、普通複数の未知パラメータが含まれている。 R_n が与えられた場合の未知パラメータベクトル θ の対数尤度は下記 (2.13) 式で与えられる。

$$(2.13) \quad l(\theta|y) = \sum_{n=1}^N \log(p(R_n|\Psi_{n-1})).$$

この対数尤度を疑似ニュートン法 (DFP 公式や BFGS 公式) 等の数値的最適化アルゴリズムを用いて、パラメータ θ に関して最大化することによりパラメータ θ の最尤推定量 $\hat{\theta}$ を求めることが出来る [13]。

また、本稿ではモデル選択の基準として、下記式で定義される AIC (Akaike's Information Criterion) を用いる [1, 21]。

$$\begin{aligned}
 AIC &= -2l(\hat{\theta}) + 2m \\
 (2.14) \quad &= -2(\text{最大対数尤度}) + 2(\text{パラメータ数}).
 \end{aligned}$$

3 実証分析

本節では、上記で提案したマルコフ切換モデルをマーケティングにおいて活用される販売データである POS データへ適用することによって、本モデルの有用性の検証を行う。

具体的には、1 節に示した問題意識に基づき、過去の時点における非価格プロモーション実施の有無の推定を試みた。次に、たまたまこのデータセットのみ POS データに併せて調査により取得された非価格プロモーション実施状況のデータを用い、本モデルにより推定された非価格プロモーションの実施の有無が実際の実施状況とどの程度適合しているかの検証を行った。さらに本稿で提案した手法の評価を視野に入れて、教師データが存在すると仮定した非線形判別分析との比較も行った。

なお、スーパーマーケットで販売される商品の売上時系列を取り扱う場合、データの季節性を考慮することが通常は必要である。アイスクリーム、チョコレート、ビールなどの売り

上げを考えていただければ明らかであろう。しかしながら、今回分析対象としたインスタントコーヒーでは、季節性は売上にほとんど影響しないことが実務上知られている。そのため、今回の解析においては季節性は考慮していない。ただし、より一般的なPOSデータを用いた解析では前述のように季節性を考慮することが必要である。季節性の処理に関しては大別して2つの方策が考えられる。1つ目は、移動平均法を用いてモデルを当てはめる前に季節性による影響を事前にデータから除去するものである。ただし、この手法は、全ての時系列に対して、既存の移動平均フィルタで処理するために、様々な因子による効果を画一的にデータからあらかじめ除去してしまうおそれがあり、情報抽出の観点から望ましくない。2つ目は、モデル内で季節性を取り込んでモデル化するものである。非定常時系列解析の立場からは、この方策が妥当であると考えられる。しかしながら、そのモデルの推定は複雑なものになるため[13, 14]、本研究の目的を鑑み、本提案モデルでは季節性は考慮していない。季節性を取り込んだモデル化については今後の課題とする。

3.1 データ

関東地方にあるスーパーマーケットの週別POSデータを用いてモデル検証を行った。対象ブランドはインスタントコーヒーの一つのブランドである。

POSデータからは、週別の販売点数、販売金額及び来店客数が獲得できる。以降の分析では、以下の式により算定される変数を用いて分析を行った。

$$(3.1) \quad y_n = \frac{UNIT_n \times 1000}{CUSTOMER_n}.$$

(3.1)式は来店客1000人当たりの販売点数を示し、通常点数PI(Purchase Incidence)と呼ばれる。

価格に関しては以降でその分布に対してベータ分布へ当てはめるために、下記(3.2)式に示す正規化を行った。

$$(3.2) \quad x_n = 0.001 + \frac{(PRICE_n - PRICE_{min})}{(PRICE_{max} - PRICE_{min})}(0.999 - 0.001).$$

上記式において、 $UNIT_n$ 、 $CUSTOMER_n$ 及び $PRICE_n$ はそれぞれ、第n週における販売点数、来店客数、売価を示している。また、 $PRICE_{max}$ 及び $PRICE_{min}$ はデータ期間中の当該ブランドにおける最大売価と最低売価をそれぞれ示している。さらに本モデルの検証のために、スーパーマーケットの店頭で調査員により毎日取得されている非価格プロモーション実施状況のデータを併せて用いた。

Fig.1は、ブランドAの点数PIと正規化売価及び調査により取得されている非価格プロモーションの時系列を示している。

Fig.2は、非価格プロモーション実施状況の先週と今週の組み合わせで、(非実施、非実施)、(非実施、実施)、(実施、非実施)、(実施、実施)(以降ではそれぞれ状態1、状態2、状態3、状態4とする)の4状態毎の、正規化売価に対する非価格プロモーション実施週数比率を示している。図から読みとれるように、状態1と状態3は価格帯が高い方へ、また状態2と状態4は価格帯が低い方へ歪んでいる。

Fig.3は、分析対象全データの正規化売価と点数PIの散布図を示している。また、Fig.4は、Fig.3と同じデータを各状態毎に分割して図示したものである。今週非価格プロモーションを実施していない場合(Fig.4上段)、点数PIの分散が小さく、大部分が正規化売価0.5~1.0に分布している。逆に、今週非価格プロモーションを実施している場合(Fig.4下段)、点数PIの分散が大きく、大部分が正規化売価0.2~0.6に分布している。このように非価格プロモーションの実施状況に応じて、点数PI及び正規化売価の分布状況に違いが生じる。

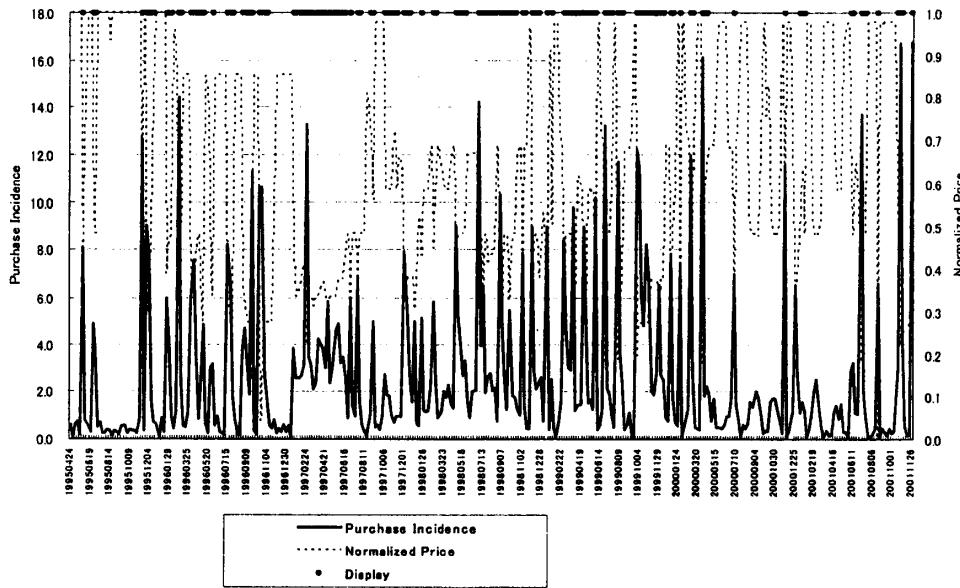


Fig. 1: Time Series of Purchase Incidence and Normalized Price

3.2 非価格プロモーション実施の有無を推定するためのモデル

Table 1: Features of Models

Number of Regime	Regime	Assumption of State	Number of Parameter		
			Distribution	Regression	2-Variate
2state	$S_n = 1$	Regime1 $\phi_n = 0$	6	8	12
	$S_n = 2$	Regime2 $\phi_n = 1$			
3state	$S_n = 1$	Regime1 $\phi_n = 0$	9	12	18
	$S_n = 2$	Regime2 $\phi_{n-1} = 0$			
	$S_n = 3$	Regime3 $\phi_{n-1} = 1$			
4state	$S_n = 1$	Regime1 $\phi_{n-1} = 0$	12	16	24
	$S_n = 2$	Regime2 $\phi_{n-1} = 0$			
	$S_n = 3$	Regime3 $\phi_{n-1} = 1$			
	$S_n = 4$	Regime4 $\phi_{n-1} = 1$			

本節では、2章で提案したマルコフ切換モデルを3.1節で詳説したデータに適用し、Table 1に示す9つのモデルの推定及び検証を行った。ただし、非価格プロモーション実施の有無に関する情報は用いていない。

非価格プロモーション実施の有無推定のために想定した観測モデルと状態数毎に仮定したシステムモデルの説明を行う。2章で定義したMSDモデル、MSRモデル、MSBモデルの状態空間モデル表現における観測モデルは、具体的にそれぞれ下記のモデルを仮定する。なお、下記表中 ϕ_n は n 週目の非価格プロモーション実施の有無を示すダミー変数である。

(観測モデル：MSDモデル)

各状態毎に、正規化売価 x_n の分布が下記に示すようにベータ分布に従うと仮定する。

$$(3.3) \quad g(x_n | S_n = i, \psi_{n-1}^x) = g_i(x_n) = \frac{x_n^{\gamma_i - 1} (1 - x_n)^{\delta_i - 1}}{B(\gamma_i, \delta_i)}, \quad i = 1, \dots, k.$$

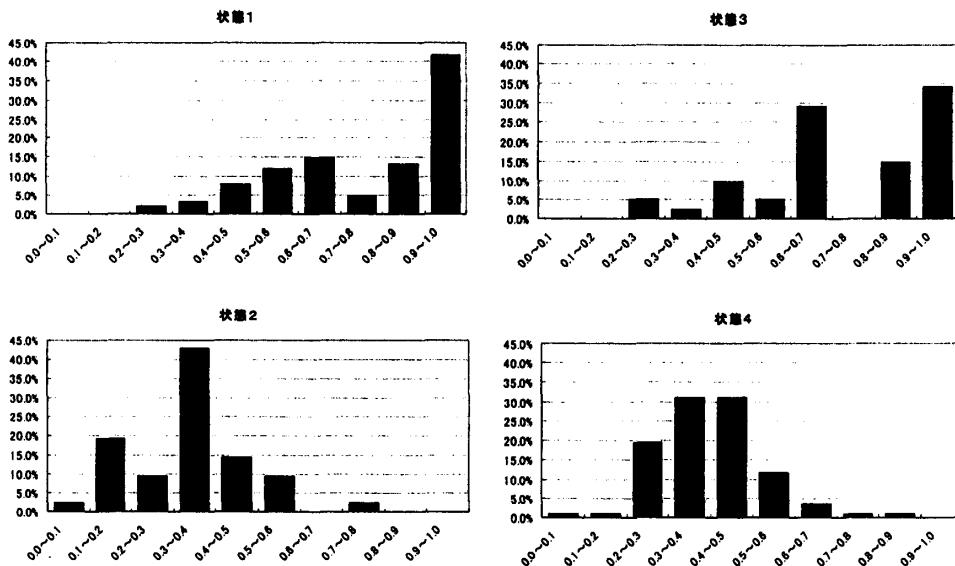


Fig. 2: Histogram of Normalized Price (Every Regime)

ここで, $B(\gamma_i, \delta_i)$ はベータ関数を示すものとする.

(観測モデル: MSRモデル)

各状態毎に点数 PI は、販売時の当該商品の正規化売価に影響されると仮定し、下記式の通常の回帰モデルに従うと仮定する。

$$(3.4) \quad f(y_h|S_n = i, \psi_{n-1}^y, x_n) = f_i(y_h|\mu_i) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left\{-\frac{1}{2\sigma_i^2}(y_h - \mu_i)^2\right\}, i = 1, \dots, k.$$

さらに、 μ_i は下記の線形結合により与えられる。

$$(3.5) \quad \mu_i = \beta_{i0} + \beta_{i1}x_n.$$

(観測モデル: MSBモデル)

各状態毎に、正規化売価と点数 PI の 2 変量時系列 (x_n, y_h) の分布が下記の分布に従うものと仮定する。

$$(3.6) \quad \begin{aligned} h(x_n, y_h|S_n = i, \psi_{n-1}^{xy}) &= f(y_h|S_n = i, \psi_{n-1}^y, x_n)g(x_n|S_n = i, \psi_{n-1}^{xy}) = f_i(y_h|\mu_i)g(x_n) \\ &= \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left\{-\frac{1}{2\sigma_i^2}(y_h - \mu_i)^2\right\} \frac{x_n^{\gamma_i-1}(1-x_n)^{\delta_i-1}}{B(\gamma_i, \delta_i)}, i = 1, \dots, k. \end{aligned}$$

(3.6) 式で定義されるMSBモデルは、(3.3), (3.4) 式に示す、MSDモデルとMSRモデルの分布の積によって規定されるモデルである。

さらに、下記でシステムモデルを説明する。本稿では、Table 1 に示したように 3 つの状態数（2 状態、3 状態、4 状態）を仮定して分析を進めている。状態数毎のそれぞれのシステムモデルは具体的に下記のように定式化できる。

(システムモデル: 2 状態モデル)

2 状態モデルでは、下記のシステムモデルを仮定した。

$$(3.7) \quad P = \begin{pmatrix} p & 1-q \\ 1-p & q \end{pmatrix}.$$

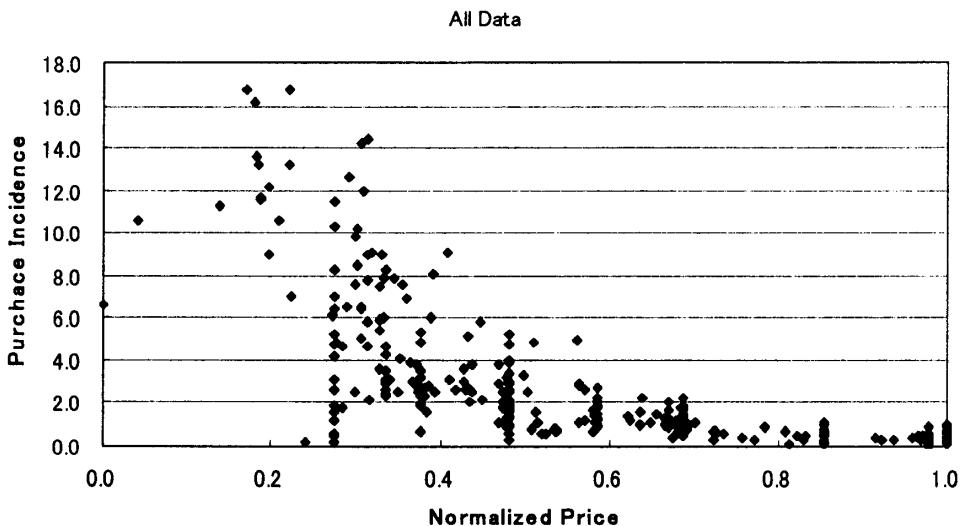


Fig. 3: Scatter Plot (All Data)

この場合、仮定した潜在的な状態は当該週において非価格プロモーションを実施していたか否かである。

(システムモデル：3状態モデル)

3状態モデルでは、下記システムモデルを仮定した。

$$(3.8) \quad P = \begin{pmatrix} p & 1-q & 1-r \\ 1-p & 0 & 0 \\ 0 & q & r \end{pmatrix}.$$

この場合仮定した潜在的な状態は、当該週において非価格プロモーションを実施していた場合、していなかった場合をさらに先週非価格プロモーションを実施していたか否かで分類した3状態である。

(システムモデル：4状態モデル)

4状態モデルでは、下記システムモデルを仮定した。

$$(3.9) \quad P = \begin{pmatrix} p & 0 & 1-r & 0 \\ 1-p & 0 & r & 0 \\ 0 & q & 0 & 1-s \\ 0 & 1-q & 0 & s \end{pmatrix}.$$

この場合、仮定した潜在的な状態は3.1節で説明したように非価格プロモーション実施の有無の前週と今週の組み合わせで、(無, 無), (無, 有), (有, 無), (有, 有)の4状態である。

なお、具体的に(2.13)式中のパラメータ θ は、4状態MSBモデルを例とすると、 $\theta = \{\beta_{1,0}, \beta_{1,1}, \sigma_1^2, \dots, \beta_{4,0}, \beta_{4,1}, \sigma_4^2, \gamma_1, \delta_1, \dots, \gamma_4, \delta_4, p, q, r, s\}$ となる。

3.3 モデル推定結果

Table 2は、各モデルの良さをAICにより比較した結果である。ただし、MSD, MSR, MSBの異なるモデル間では尤度を構成しているデータが異なっているため、計っている尤度がそもそも違う。しかしながら、表の縦のモデル比較はAICによって可能である。いずれのモ

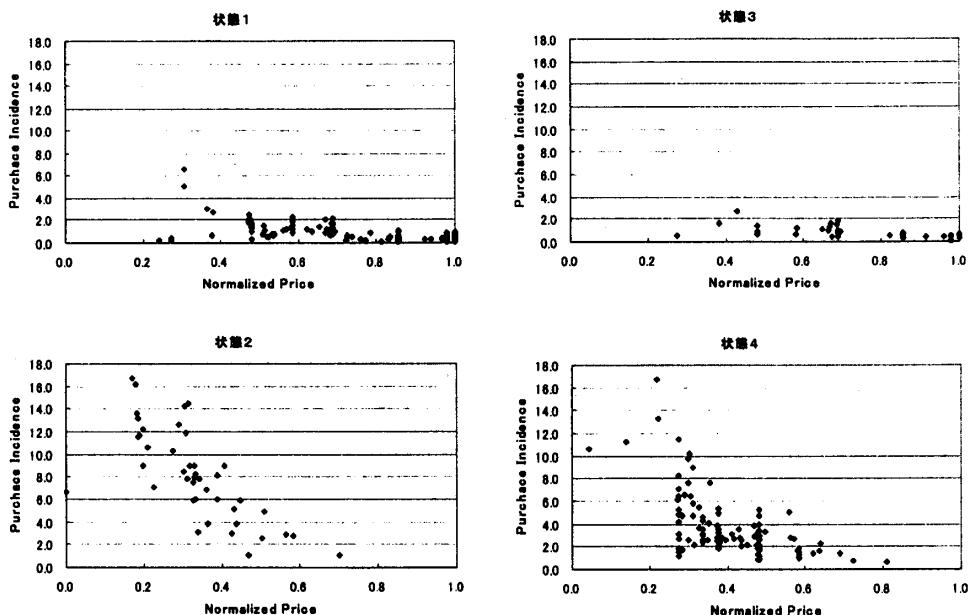


Fig. 4: Scatter Plot (Every Regime)

モデルにおいても、仮定する状態数を多くするとモデルの改善がみられ、その結果4状態モデルの当てはまりが良くなっている。これは、仮定する状態数を増やすことにより、時系列の状態を的確に捉えることが可能になったことを示すものである。

Table 2: Calculation Result of log-likelihood and AIC

Number of State	MSD		MSR		MSB	
	log-likelihood	AIC	log-likelihood	AIC	log-likelihood	AIC
2	173.49	-334.97	-572.46	1160.92	-360.95	745.91
3	175.98	-333.95	-560.31	1144.61	-325.39	686.78
4	182.31	-340.62	-545.09	1122.18	-312.08	672.17

3.4 非価格プロモーション実施の推定値と観測値の比較

前節で示したように、各モデル毎にモデルの良さを AIC によって比較すると、いずれのモデルにおいても4状態モデルの当てはまりが最も良い。そのため、本節では4状態のMSDモデル、MSRモデル、MSBモデルの3つのモデルそれぞれの平滑化確率を用いて、非価格プロモーション実施の有無（実績値）と非価格プロモーション実施の有無（推定値）を比較し、本モデルの判別モデルとしての良さを検証した。4状態モデルで推定されている平滑化確率は、非価格プロモーション実施の前週と今週の組み合わせた変数に対応して定義されている。我々の興味は、今週の非価格プロモーション実施の有無を推定することであるため、本検証においては下記式の示す平滑化確率の周辺確率を算定し、判別を行っている。

（周辺平滑化確率）

$$(3.10) \quad Pr(\phi_n = 0 | \psi_N) = Pr(S_n = 1 | \psi_N) + Pr(S_n = 3 | \psi_N),$$

$$(3.11) \quad Pr(\phi_n = 1 | \psi_N) = Pr(S_n = 2 | \psi_N) + Pr(S_n = 4 | \psi_N).$$

ここで、上記(3.10)式が今週の非価格プロモーション非実施の周辺平滑化確率を示し、逆に(3.11)式は非価格プロモーション実施の確率を示している。Fig.5には、非価格プロモーション実施の周辺平滑化確率と実績をモデル毎に示している。図から判定すると、MSDモデルは同じ状態を継続しており、MSRモデルは切換が数多く発生している。またMSBモデルは、MSDモデルとMSRモデルの中間的な様相になっている。

下記(3.12)式により、モデル毎の周辺平滑化確率を用いて非価格プロモーション実施の有無の判別を行った。さらにこの結果を用い、(3.13)式で実績値と推定値の一致率を算定することで、各モデル毎の判別の良さを検証した。

$$(3.12) \quad D_n = \begin{cases} 1, & \text{if } P(\phi_n = 1 | \psi_N) \geq 0.5, (\text{実施:Execution}), \\ 0, & \text{if } P(\phi_n = 1 | \psi_N) < 0.5, (\text{非実施:Non-Execution}). \end{cases}$$

D_n は時点 n での非価格プロモーション実施の有無の判別結果を示している。

$$(3.13) \quad RC = \frac{\sum_{n=1}^N I(O_n = D_n)}{N}.$$

上記(3.13)式内において、 O_n は非価格プロモーション実施の有無の実績値（有：1、無：0）、 N はサンプル数。また、 $I(\cdot)$ は下記に示すような定義関数である。

$$(3.14) \quad I(\cdot) = \begin{cases} 1, & \text{if } O_n = D_n, \\ 0, & \text{if } O_n \neq D_n. \end{cases}$$

上記に示した方法により、調査により獲得できている非価格プロモーションの実施状況のデータと、モデルに基づく推定結果との比較をモデル毎に行い、誤分類サンプルを識別した上で散布図を記述したものを Fig.6 に示す。Fig.6 の上段は MSD モデルの、中段は MSR モデルの、下段は MSB モデルの散布図を示している。MSD モデルでは、正規化売価 0.5~0.7 付近のサンプルにおいて、非実施サンプルを実施サンプルと誤分類しているケースが多い。また、正規化売価 0.2~0.5 の区間においても、非実施サンプルを実施サンプルと誤分類しているものがみられる。MSR モデルでは、MSD モデルで誤分類が多く発生した正規化売価 0.5~0.7 の区間の判別は改善しているが、点数 PI が 2~4 の間の非価格プロモーション実施サンプルの誤分類が多くなっている。MSB モデルによる判別は、MSD モデルと比較すると非価格プロモーションを実施していないサンプルが、また MSR モデルと比較すると非価格プロモーション実施時のサンプルの判別率が向上しており、全体としては最も判別率が高いモデルである。誤分類率をモデル毎に整理したものを、Table 3 に示す。そこから判断できるように、MSB モデルの判別率が他の 2 つのモデルと比較して高い。すなわち、本研究対象データにより目的である過去の時点における非価格プロモーションの実施状況を判別する際に、最も正確に判別できるモデルは MSB モデルである。これは、上記で推定結果を用いて言及したように、MSB モデルが MSD モデル及び MSR モデルそれぞれの短所を克服したモデルであるということから、妥当な結果であると判断できる。

次に、上記の検証により最良のモデルであると判定された 4 状態 MSB モデルの回帰部分の推定結果を用いて、残差の状況を確認する。

提案したモデルでは、状態毎に異なる回帰係数及び誤差分散を仮定している。すなわち、それらのパラメータは状態の推移に従属して時点毎に変化する。また、各時点での観測値の確率密度は(2.10)式に示すように、状態の予測分布で重み付けされた混合分布の形状をしているため、一般には通常の回帰モデルのように残差を定義できない。しかし、今回の場合、推定された状態確率を時点毎にみると、概ね一つの状態が支配的であることが検証より分かって

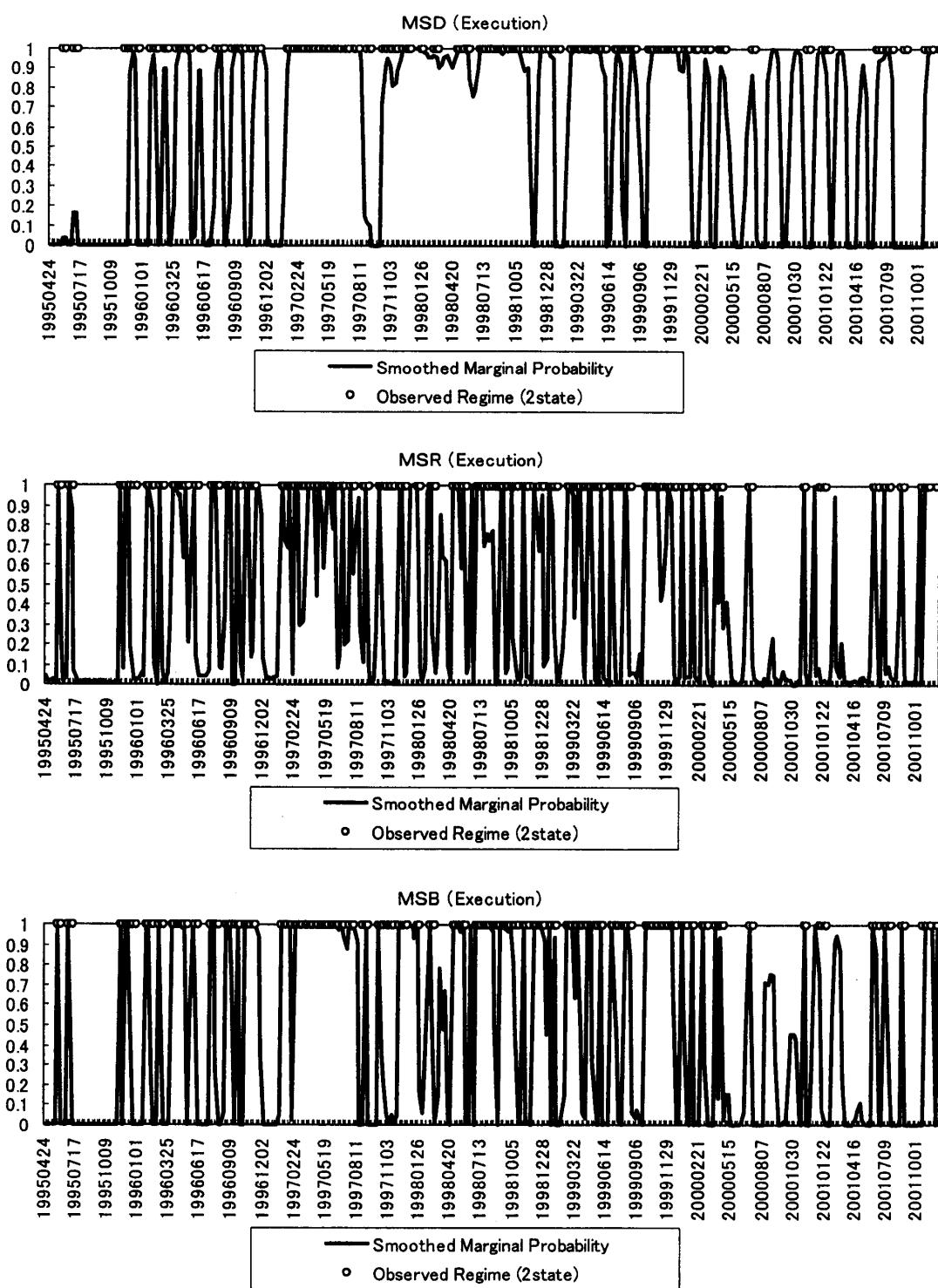


Fig. 5: Marginal Smoothed Probability

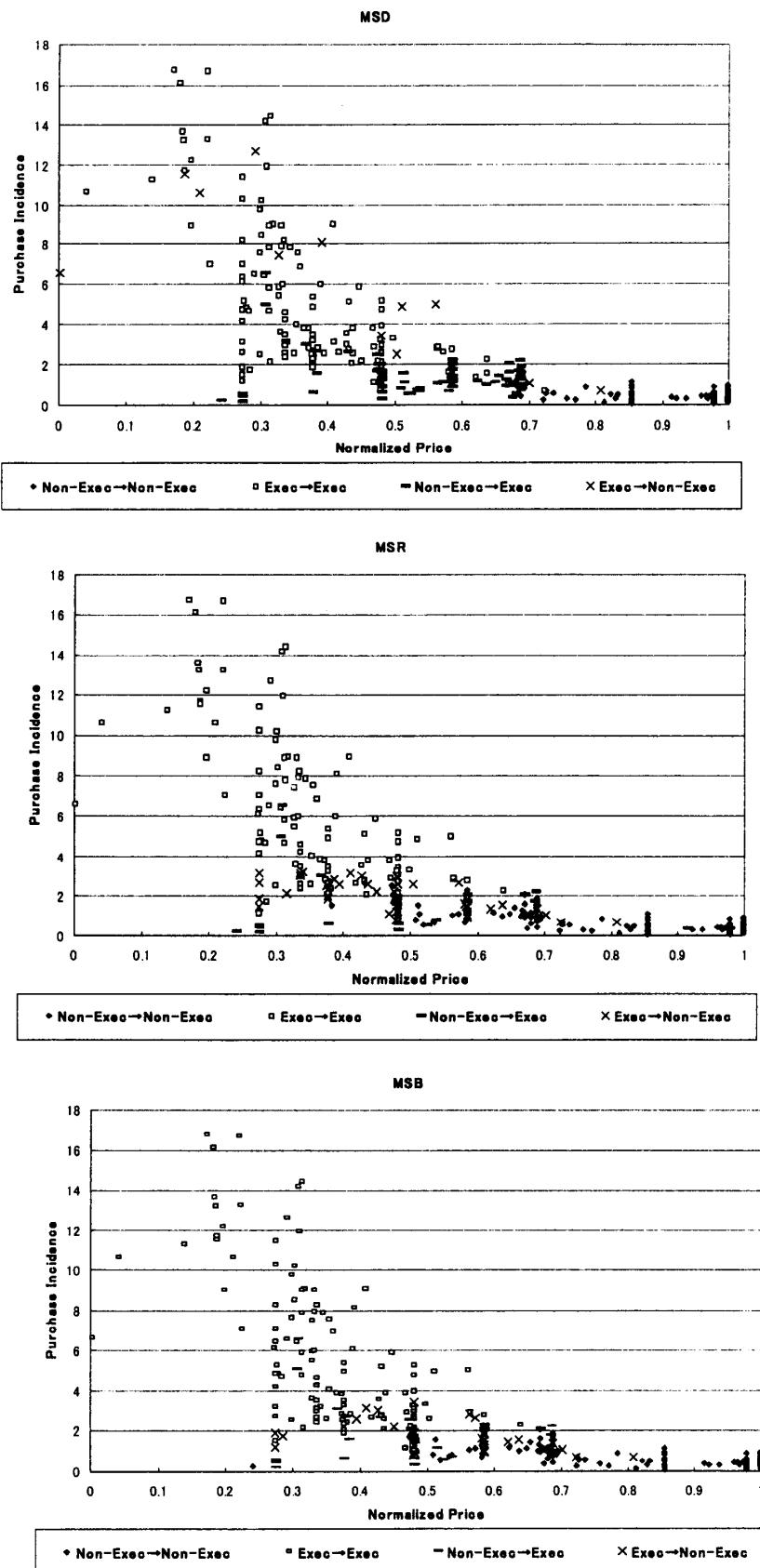


Fig. 6: Scatter Plot of Discrimination Result

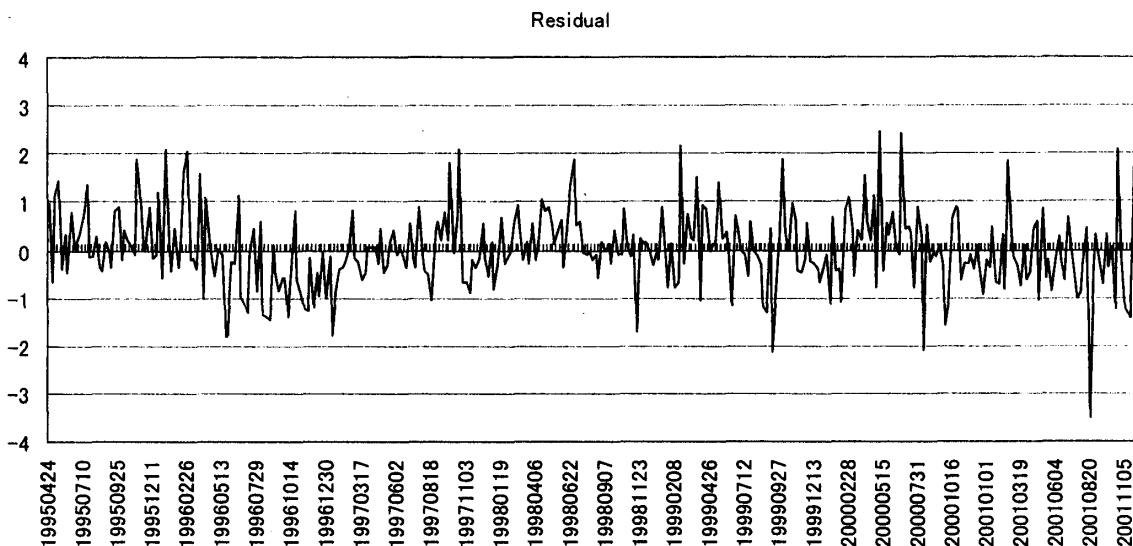


Fig. 7: Time Series of Residual

Table 3: Misclassification Rate

Model	Variable	Non-execution	Execution	Total
M S D	Misclassification Rate	0.38	0.08	0.25
	Number of Misclassification Sample	73	13	86
M S R	Misclassification Rate	0.11	0.31	0.20
	Number of Misclassification Sample	21	48	69
M S B	Misclassification Rate	0.16	0.20	0.18
	Number of Misclassification Sample	30	31	61

いるため、以降では下記(3.15)式により残差を定義し、その系列を用いて残差の検証を行う。

$$(3.15) \quad \hat{\epsilon}_{n|J} = \frac{y_n - \beta_{\hat{i}_n 0} - \beta_{\hat{i}_n 1} x_n}{\sigma_{\hat{i}_n}}.$$

ただし、 $\sigma_{\hat{i}_n}$ は推定された状態毎の誤差分散の平方根である。また、 \hat{i}_n は下記式により規定される。

$$(3.16) \quad \hat{i}_n = \operatorname{argmax}_i Pr(S_n = i | \psi_J).$$

ここで、(3.15)及び(3.16)式は J の値により、予測($J = n - 1$)、フィルタ($J = n$)及び平滑化($J = N$)の3通りのケースが考えられる。以降では、 $J = N$ の平滑化の推定結果を用いて検証を行う。

Fig.7は(3.15)式により算定した残差の系列である。この系列からダービンワトソン比(DW)*を算定すると1.79である。Fig.8はFig.7と同じデータを状態毎に分割して図示したものである。状態毎の残差は時間とともに均一に変化している。またFig.9は残差と正規化売価の散布状況を状態毎に図示したものである。この図からも各状態とも分散均一性が確認

*ダービンワトソン比は、 $0 < DW < 2$ の場合正の系列相関があり、 $2 < DW < 4$ の場合負の系列相関と判断する。一般にダービンワトソン比は2に近づくほど良い(系列相関がない)とされる。

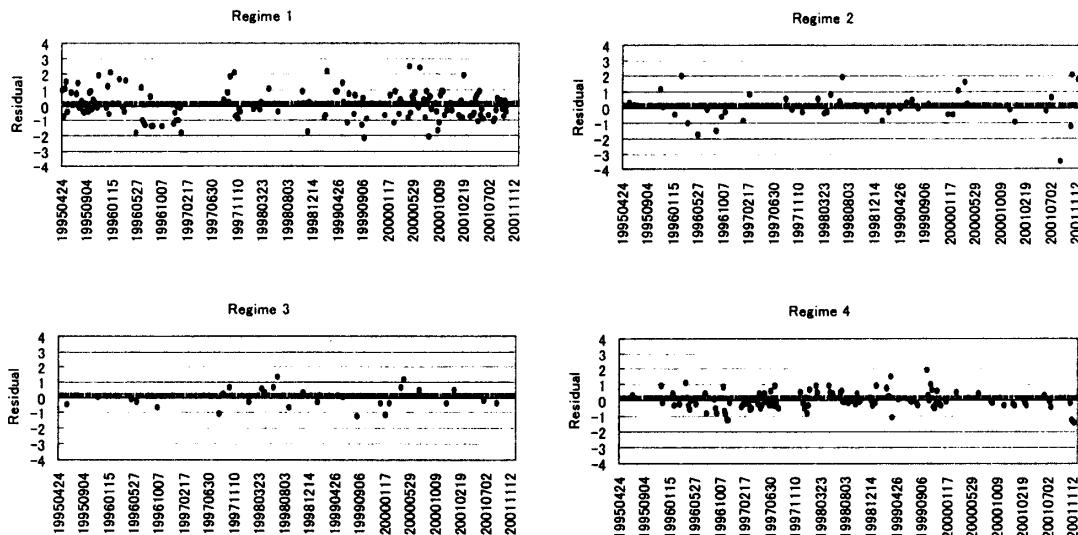


Fig. 8: Time Series of Residual (Every Regime)

できる。さらに、状態毎に残差の歪度 (Skewness) と尖度 (Kurtosis) を算定したものを Table 4 に示す。そこから誤差項の正規性は確認できる*。

Table 4: Skewness and Kurtosis

Regime	Skewness	Kurtosis
Regime 1	0.3292	0.0872
Regime 2	-0.5460	2.3964
Regime 3	0.0562	-0.1158
Regime 4	0.3015	1.2822

3.5 非線形判別分析による非価格プロモーション実施の有無の判別

本小節ではマルコフ切換モデルによる推定の評価を別の視点から行うために、教師データが存在すると仮定した問題設定の場合の非線形判別分析で、非価格プロモーション実施の有無がどの程度判別されるかに関して検証を行った。その際用いる変数は点数 PI と正規化売価である。検証は 3 つのケースを設定して行った。ケース 1 及び 2 は、マルコフ切換モデルによる判別結果と比較するために、分析対象データの時系列構造を担保した形で 2 つに分割し検証している。またケース 3 は、分析対象データをランダムに 2 つに分割し検証している。そのため、データの時系列構造は無視しているので、MSB モデルには適用できない。このケースはランダムな教師データを用いた場合、RBF の判別精度がどの程度なのかを確認するために行った。ケース 1 は、Fig.10 左側に示す Data1 を非線形判別分析の学習用のデータとして使用した。次に、非線形判別分析の学習結果より Fig.10 右側に示す Data2 の判別分析を行う。同様に Data2 に対して、MSB モデルを適用し非価格プロモーション実施の有無の判別を行う。ケース 2 は、ケース 1 の学習用及び評価用のデータを逆に使用し、Data2 を RBF の学習用に、Data1 を RBF の評価と MSB モデルの推定に用いるケースである。この 2 ケ

* $|Skewness| < 2, |Kurtosis| < 4$ であれば、正規性の仮定を満たすと考えられる。

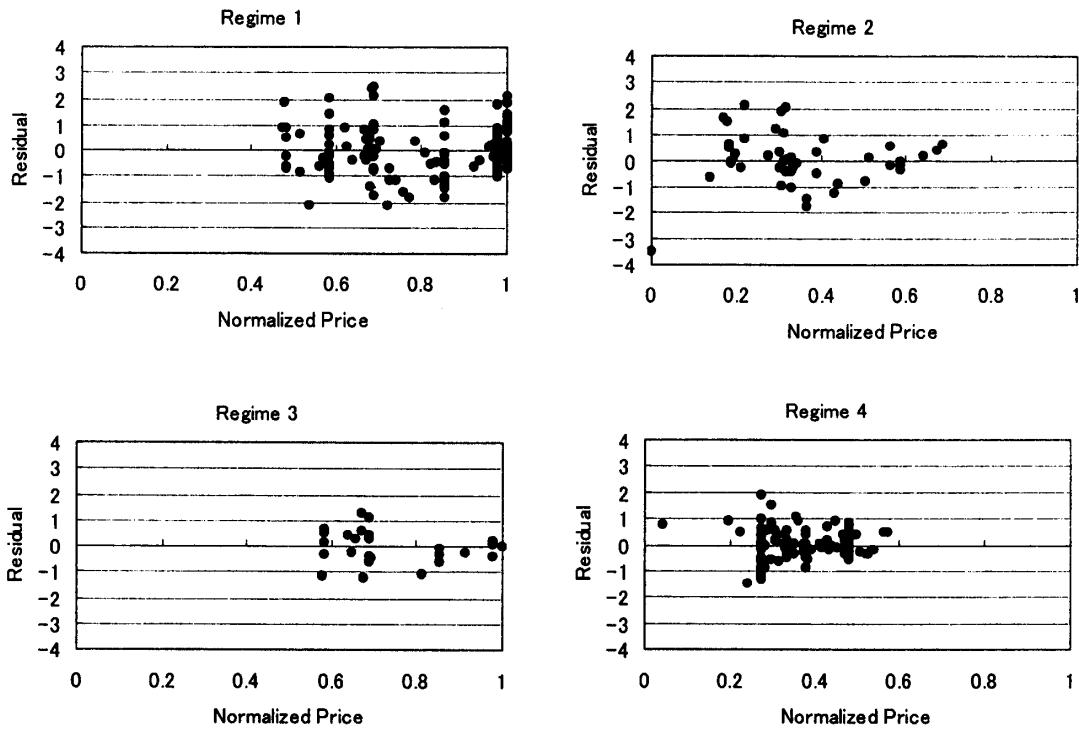


Fig. 9: Scatter Plot (Normalized Price vs Residual: Every Regime)

スでは最終的に、非価格プロモーション実施の有無の実績と推定結果から誤分類率を算定し両手法の性能比較を行った。ケース3では、Fig.11の左側に示すData3を学習用に、右側に示すData4を評価用データとしてRBFの判別精度の検証を行っている。

非線形判別分析には、最近パターン認識などの判別分析で数多くの研究がなされているサポート・ベクター・マシンが視野に入るべきであるが、今回は階層型ニューラルネットワークモデルの一種である動径基底ネットワーク (Radial Basis Function Networks, 以降RBF) を用いた。なお、サポートベクターマシンに関する詳細は、[2]を参照のこと。

RBFネットワークは、Fig.12に示すように、入力層と1層からなる中間層及び出力層の3層から構成され、学習データ $\{X_i \in R^d | i = 1, \dots, p\}$ と群のラベル $G \in \{1, \dots, L\}$ に関する N 個の観測データ $\{(x_\alpha, g_\alpha); \alpha = 1, \dots, N\}$ に基づいて L 群非線形判別方程式が定式化される。まずははじめに関数 $F_k(x)$ を次のように与える。

$$(3.17) \quad F_k(x) = \sum_{i=1}^M \omega_{ki} \psi_i(x - x_i), \quad k = 1, \dots, L-1.$$

このとき $x_i \in R^d$ をRBFの中心という。中間層は M 個の素子からなりその活性化関数を $\psi_i(x - x_i)$ とする。また本稿では活性化関数として、下記に示すガウス基底関数を用いている。

$$(3.18) \quad \psi_i(r_i) = \exp\left\{-\frac{r_i^2}{2\sigma^2}\right\}, \quad \sigma > 0, r_i \geq 0.$$

ここで、 $r_i \equiv \|x - x_i\|$ とする。次に、データ x_α が観測されたとき、それが第 k 群である事後確率を $Pr(g_\alpha = k | x_\alpha)$, $k = 1, \dots, L$ とする。

このとき次のような非線形判別関数を考える。

$$(3.19) \quad \log \frac{Pr(g_\alpha = k | x_\alpha)}{Pr(g_\alpha = L | x_\alpha)} = F_k(x), \quad k = 1, \dots, L-1.$$

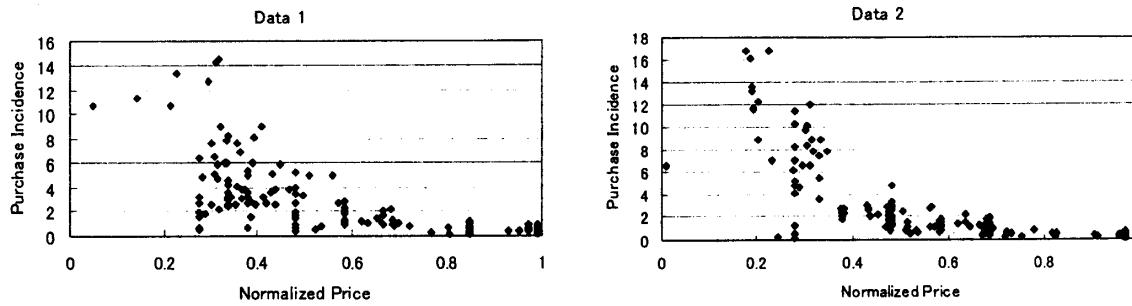


Fig. 10: Scatter Plot of Training and Evaluation Data (Case-1 and Case-2)

対数オッズ比に非線形性を仮定した(3.17)式のモデルは、

$$\Pr(g_\alpha = k|x_\alpha) = \frac{\exp\{F_k(x)\}}{1 + \sum_{j=1}^{L-1} \exp\{F_j(x)\}},$$

$$\Pr(g_\alpha = L|x_\alpha) = 1 - \sum_{k=1}^{L-1} \Pr(g_\alpha = k|x_\alpha) = \frac{1}{1 + \sum_{j=1}^{L-1} \exp\{F_j(x)\}},$$

と同値であり、明らかに $\sum_{k=1}^L \Pr(g_\alpha = k|x_\alpha) = 1$ が成り立つ。このデータ x_α が第 k 群から観測されたとする確率 $\Pr(g_\alpha = k|x_\alpha)$ を用いてサンプルの判別を行う。今回の場合のように非価格プロモーションを実施していたか否かの 2 群の場合は、前述した(3.12)式と同様に判別される。なお、RBF ネットワークに関する研究としては、[22] の優れたサーベイ報告がある。RBF ネットワークの詳細に関しては、当該論文を参照のこと。

今回は(3.1), (3.2)式で算定される点数 PI 及び正規化売価を入力層とし、非価格プロモーション実施の有無(ダミー変数)を出力層とした。

Fig.13, 14 には、ケース 1, 2 それぞれの RBF ネットワークと MSB モデルによる判別結果を図示している。Fig.15 には、ケース 3 における RBF ネットワークによる判別結果を図示している。また、Table 5 はモデル毎に誤分類率を算定した結果である。ケース 1 及び 2 において、MSB モデルによる誤分類率は、教師データを必要とする RBF ネットワークの誤分類率とほぼ同程度である。この利点はモデルへ動的な特性を明示的に取り込んでいる点によってもたらされている。また、各ケース間で RBF ネットワークによる判別力は同程度である。本稿で提案したモデルは教師データが存在しない状態でも、的確にサンプルの状態を捉えることが可能であるという点で非常に有用であるといえる。

3.6 実証分析からの考察

今回分析対象とした商品における非価格プロモーション実施の有無の実績値と推定値の一致率は、今回最も精度高く判別できた MSB モデルの場合で、0.82 程度であった。また、教師データが存在しない本提案モデルは、教師データが存在する下で実施される非線形判別分析とほぼ同程度の判別能力があることが示された。今後さらに一致率の向上を図るためにには、いくつかの方法が考えられる。ここでは、3 つの可能性について言及する。

1 つは、仮定する状態毎のモデルに関するものである。今回は、点数 PI と正規化売価の関係は通常の線形回帰モデルを仮定した。しかし、前掲の Fig.4 の右下のグラフのように状態毎

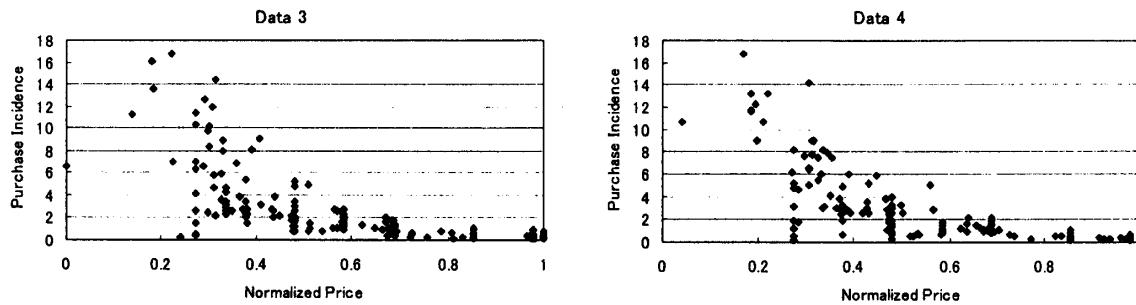


Fig. 11: Scatter Plot of Training and Evaluation Data (Case-3)

Table 5: Misclassification Rate of RBF and MSB

Case	Model	Variable	Non-Execution	Execution	Total
Case-1	RBF	Misclassification Rate	0.15	0.19	0.17
		# Misclassification Samples	15	14	29
Case-1	MSB	Misclassification Rate	0.15	0.22	0.18
		# Misclassification Samples	15	16	31
Case-2	RBF	Misclassification Rate	0.12	0.12	0.12
		# Misclassification Samples	11	10	21
Case-2	MSB	Misclassification Rate	0.18	0.17	0.18
		# Misclassification Samples	16	14	30
Case-3	RBF	Misclassification Rate	0.12	0.15	0.13
		# Misclassification Samples	12	11	23
Case-3	MSB	Misclassification Rate	-	-	-
		# Misclassification Samples	-	-	-

にみるとその関係に非線形の挙動を示す部分がある。そのため、非線形性を的確に捉えるモデルを適用することで、一致率の向上が期待できる。

2つ目は、競合商品の影響を考慮したマルコフ切換多変量モデルを構築することである。通常、スーパーマーケットで販売される商品においては、複数の商品間で競合状況にあると考えるのが普通である。ある商品の売上に着目すると、競合商品が数多く売れた際にはその商品の売上は低下し、またその逆も成り立つ。そのような状況を的確に捉え、自商品の価格変動に加えて、競合商品の販売量の変動、価格の変動を取り込んだ形式の多変量モデルとして観測モデルを定式化するのである。ただし、本アプローチを採用した場合には、非常に数多くのパラメータを含むことになる。そのため、パラメータを減少させるかあるいは数多くのパラメータでも安定して推定できる手法を採用する等の工夫が必要となる。

3つ目は、今回使用したデータの期間の影響が考えられる。今回はPOSデータに提案モデルを適用するにあたり、6年間程度の週別のPOSデータを使用した。そのため、カテゴリー内の商品間の競争状況の変化などにより、実際には構造自体の変化が生じている可能性がある。この要因を考慮し一致率を向上させるためには、本提案モデルの固定パラメータを時変パラメータとして定式化し、市場の変化に反応して変化するようにすることが必要である。

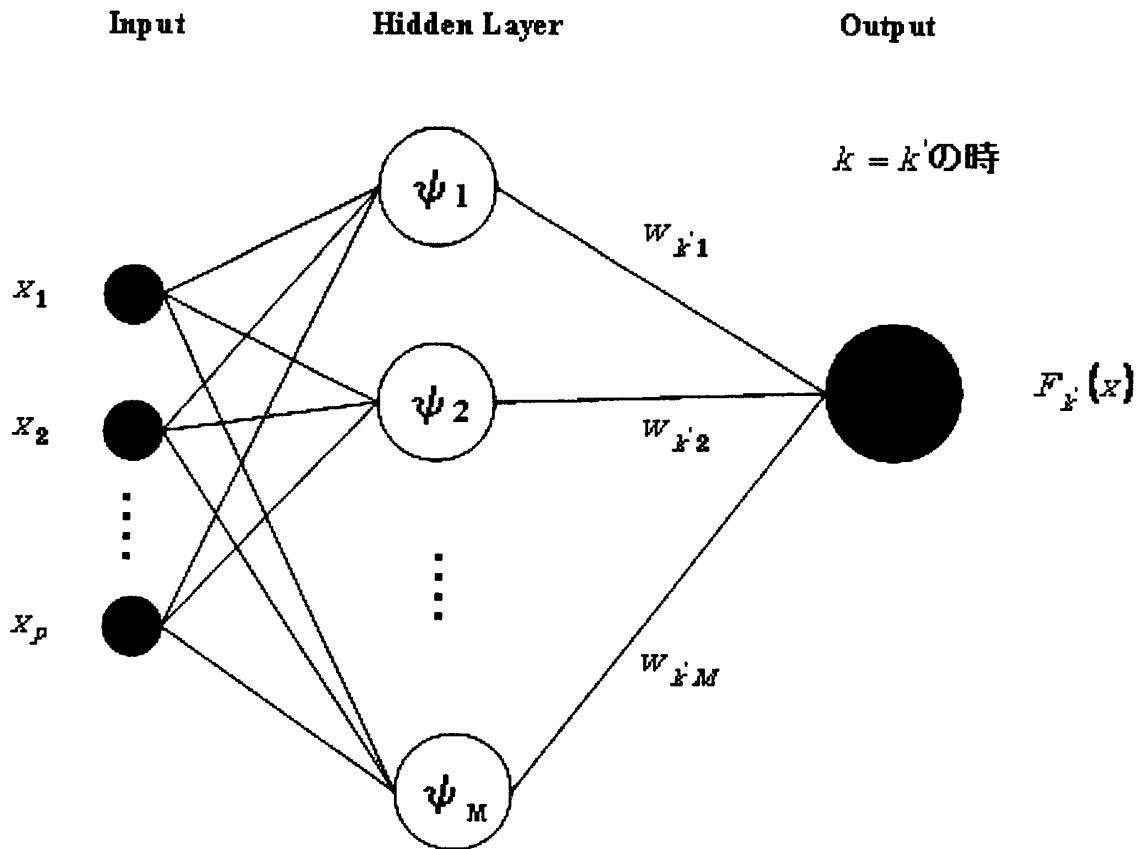


Fig. 12: RBF Network

4 数値実験

本節では、数値実験を通して提案した方法の特徴・有効性について検討する。2種類の数値実験を行った。1つ目はマルコフ切換の動的特性による、提案モデルの判別精度の変動に関するものである。2つ目は、回帰モデルの説明変数（以降、形式的に外生変数と呼ぶ）の発生機構の違いによる判別精度の変化に関するものである。数値実験においては、真のモデルを仮定し、そのモデルに従ってデータを発生させ、そのデータへMSD, MSR, MSBモデルを適用することによって検証を行った。実際には、平滑化確率を用いた状態の推定値と真のモデルから発生された状態を比較し、提案モデルの性能を検討する。以降には、その数値実験による検証結果を示す。なお、数値実験には4状態モデルを用いている。

シミュレーションでは、設定毎に300サンプルの系列を100系列ずつ発生させ、検証を行っている。なお、本節における平滑化確率を用いた判別方式は下記の通りである。

$$(4.1) \quad D_n = \operatorname{argmax}_i Pr(S_n = i | \psi_N).$$

4.1 動的特性に関する検証

本小節では、動的特性のマルコフ切換モデルの判別精度への影響について検証した結果を示す。本検証は、以下に示す4つのケースをシステムモデルの真のモデルとして仮定し、それぞれの判別精度の検証を行った。

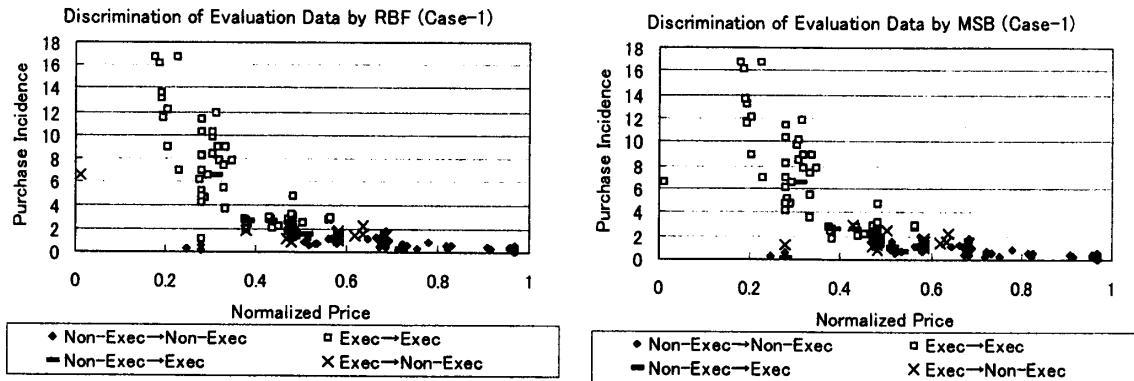


Fig. 13: Scatter Plot of Discrimination Result by RBF and MSB (Case-1)

(ケース 1 - 1)

$$(4.2) \quad P = \begin{pmatrix} 0.9 & 0 & 0.9 & 0 \\ 0.1 & 0 & 0.1 & 0 \\ 0 & 0.1 & 0 & 0.1 \\ 0 & 0.9 & 0 & 0.9 \end{pmatrix}.$$

(ケース 1 - 2)

$$(4.3) \quad P = \begin{pmatrix} 0.7 & 0 & 0.7 & 0 \\ 0.3 & 0 & 0.3 & 0 \\ 0 & 0.3 & 0 & 0.3 \\ 0 & 0.7 & 0 & 0.7 \end{pmatrix}.$$

(ケース 1 - 3)

$$(4.4) \quad P = \begin{pmatrix} 0.5 & 0 & 0.5 & 0 \\ 0.5 & 0 & 0.5 & 0 \\ 0 & 0.5 & 0 & 0.5 \\ 0 & 0.5 & 0 & 0.5 \end{pmatrix}.$$

(ケース 1 - 4)

$$(4.5) \quad P = \begin{pmatrix} 0.1 & 0 & 0.9 & 0 \\ 0.9 & 0 & 0.1 & 0 \\ 0 & 0.9 & 0 & 0.9 \\ 0 & 0.1 & 0 & 0.1 \end{pmatrix}.$$

また、観測モデルは上記に示したいずれのシステムモデルにおいても、下記に示すモデルを真のモデルとして仮定した。

(MSD モデル)

$$(4.6) \quad x_n \sim \begin{cases} B(1.79, 0.46), & \text{if } S_n = 1, \\ B(2.44, 5.07), & \text{if } S_n = 2, \\ B(2.05, 0.63), & \text{if } S_n = 3, \\ B(5.64, 8.05), & \text{if } S_n = 4. \end{cases}$$

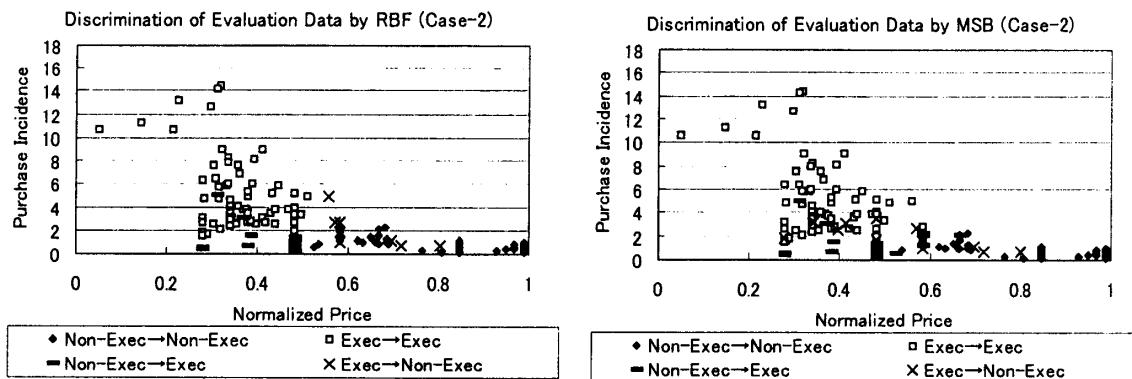


Fig. 14: Scatter Plot of Discrimination Result by RBF and MSB (Case-2)

(MSRモデル)

$$(4.7) \quad y_n \sim \begin{cases} N(2.74 - 2.46x_n, 0.44), & \text{if } S_n = 1, \\ N(15.46 - 22.12x_n, 8.34), & \text{if } S_n = 2, \\ N(1.84 - 1.44x_n, 0.19), & \text{if } S_n = 3, \\ N(9.31 - 13.77x_n, 4.80), & \text{if } S_n = 4. \end{cases}$$

(MSBモデル)

$$(4.8) \quad (y_n, x_n) \sim \begin{cases} N(2.74 - 2.46x_n, 0.44)B(1.79, 0.46), & \text{if } S_n = 1, \\ N(15.46 - 22.12x_n, 8.34)B(2.44, 5.07), & \text{if } S_n = 2, \\ N(1.84 - 1.44x_n, 0.19)B(2.05, 0.63), & \text{if } S_n = 3, \\ N(9.31 - 13.77x_n, 4.80)B(5.64, 8.05), & \text{if } S_n = 4. \end{cases}$$

Table 6 は、各システムモデルの 2 章で説明した方法で算定される状態毎の真のモデルの定常確率を示している。Fig.16 には、ケース 1-1～ケース 1-4 の順に切換が頻繁に発生し、 y 及び x の系列の変動が大きくなっている様子が示されている。Fig.17 は各ケース毎の x と y の散布図を示しており、各ケースとも Table 6 に示した定常確率の状況を的確に反映したサンプルが発生されている。

Table 6: Stationary Probabirity

Case	Regime1	Regime2	Regime3	Regime4
Case1-1	0.45	0.05	0.05	0.45
Case1-2	0.35	0.15	0.15	0.35
Case1-3	0.25	0.25	0.25	0.25
Case1-4	0.05	0.45	0.45	0.05

以降には、上記のように発生されたデータへ、MSD, MSR, MSB モデルを当てはめる。なお、検証に際してはシミュレーションで発生された既知の状態を、モデルから推定された状態と比較することにより判別精度の検証を行う。

以降での検証のために、はじめに平均誤分類率 (Average of Misclassification Rate, 以降 AMR) と MSB モデルの誤分類率が MSD, MSR よりも小さい比率 (Out-Performed Rate,

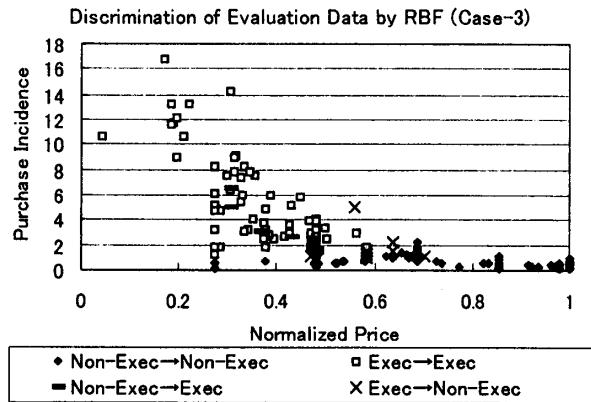


Fig. 15: Scatter Plot of Discrimination Result by RBF (Case-3)

以降OPR)を下記のように定義する.

(AMRの算定式)

$$(4.9) \quad AMR_j = \frac{\sum_{i=1}^{100} MR_i^j}{100}.$$

上記式は、各モデル毎ケース毎の平均誤分類率を示し、この値が小さいほど判別率が高いことを示す指標になる。なお、 AMR_1 、 AMR_2 、 AMR_3 はそれぞれMSDモデル、MSRモデル及びMSBモデルの平均誤分類率を示す。

(OPRの算定式)

$$(4.10) \quad OPR = \frac{\sum_{i=1}^{100} I(MR_i^3 \leq \min_{j=1}^3 MR_i^j)}{100}.$$

上記式中、 MR_i^1 、 MR_i^2 、 MR_i^3 はMSDモデル、MSRモデル及びMSBモデルの*i*番目のデータ系列の各誤判別率をそれぞれ示している。また、 $I(\cdot)$ は定義関数を示すものとする。

Table 7~10から明らかなように、いずれのケースにおいても、MSD→MSR→MSBの順に誤分類率が小さくなっている。判別モデルとしてはMSBモデルの性能が最も高い。OPRからも、ケース1~4の2ケースを除いてMSD及びMSRモデルよりMSBモデルの判別精度が高いことが分かる。

次にモデル毎に、動的特性の違いによる誤分類率を比較するといずれのモデルにおいても(ケース1~4)→(ケース1~1)→(ケース1~2)→(ケース1~3)の順に誤分類率が大きくなる傾向にある。これはTable 6に示した定常確率と関係があると判断できる。ケース1~4及びケース1~1は、前者が状態が激しく切り替わる設定で、後者が同じ状態が継続する設定である。この両者は切換の発生という点では大きな違いがあるが、定常確率からみると、前者は状態2及び3の、後者が状態1及び4の定常確率が0.45であり、ある特定の状態に時系列サンプルが属する確率が高くなっている。一方、ケース1~3では時系列サンプルはいずれの状態へも等確率で属し、AMRも他のケースに比べて大きくなっている。以上の検証内容を踏まえると、本モデルがケース1~1やケース1~4のような動的な特性を持つような現象をもつサンプルを判別する際に、より高い分類力を發揮するモデルであるということを示唆していると考えられる。

Table 7: Misclassification Rate (Case1-1)

Actual Regime	AMR						out-performed rate of MSB
	MSD		MSR		MSB		
	ave.	std.	ave.	std.	ave.	std.	
Regime1	0.1980	0.1634	0.0441	0.0366	0.0179	0.0153	0.79
Regime2	0.5053	0.2188	0.2901	0.1151	0.1552	0.0928	0.93
Regime3	0.4738	0.1863	0.6576	0.2024	0.3295	0.1275	0.86
Regime4	0.0291	0.0380	0.1219	0.0660	0.0233	0.0208	0.57
Total	0.1509	0.0676	0.1191	0.0360	0.0429	0.0159	1.00

Table 8: Misclassification Rate (Case1-2)

Actual Ragime	AMR						out-performed rate of MSB
	MSD		MSR		MSB		
	ave.	std.	ave.	std.	ave.	std.	
Regime1	0.2844	0.1315	0.1281	0.0741	0.0444	0.0302	0.89
Regime2	0.3918	0.1323	0.2770	0.0930	0.1470	0.0508	0.96
Regime3	0.4520	0.1376	0.3355	0.1141	0.2435	0.0645	0.80
Regime4	0.0797	0.0869	0.2285	0.0735	0.0777	0.0419	0.45
Total	0.2549	0.0517	0.2149	0.0417	0.1008	0.0223	1.00

Table 9: Misclassification Rate (Case1-3)

Actual Ragime	AMR						out-performed rate of MSB
	MSD		MSR		MSB		
	ave.	std.	ave.	std.	ave.	std.	
Regime1	0.3966	0.1352	0.3075	0.1494	0.0790	0.0438	0.98
Regime2	0.2879	0.0865	0.2076	0.0606	0.1186	0.0397	0.96
Regime3	0.3442	0.0903	0.1862	0.0552	0.1322	0.0458	0.87
Regime4	0.1630	0.1654	0.3006	0.0969	0.1673	0.0702	0.35
Total	0.2966	0.0567	0.2486	0.0411	0.1230	0.0236	1.00

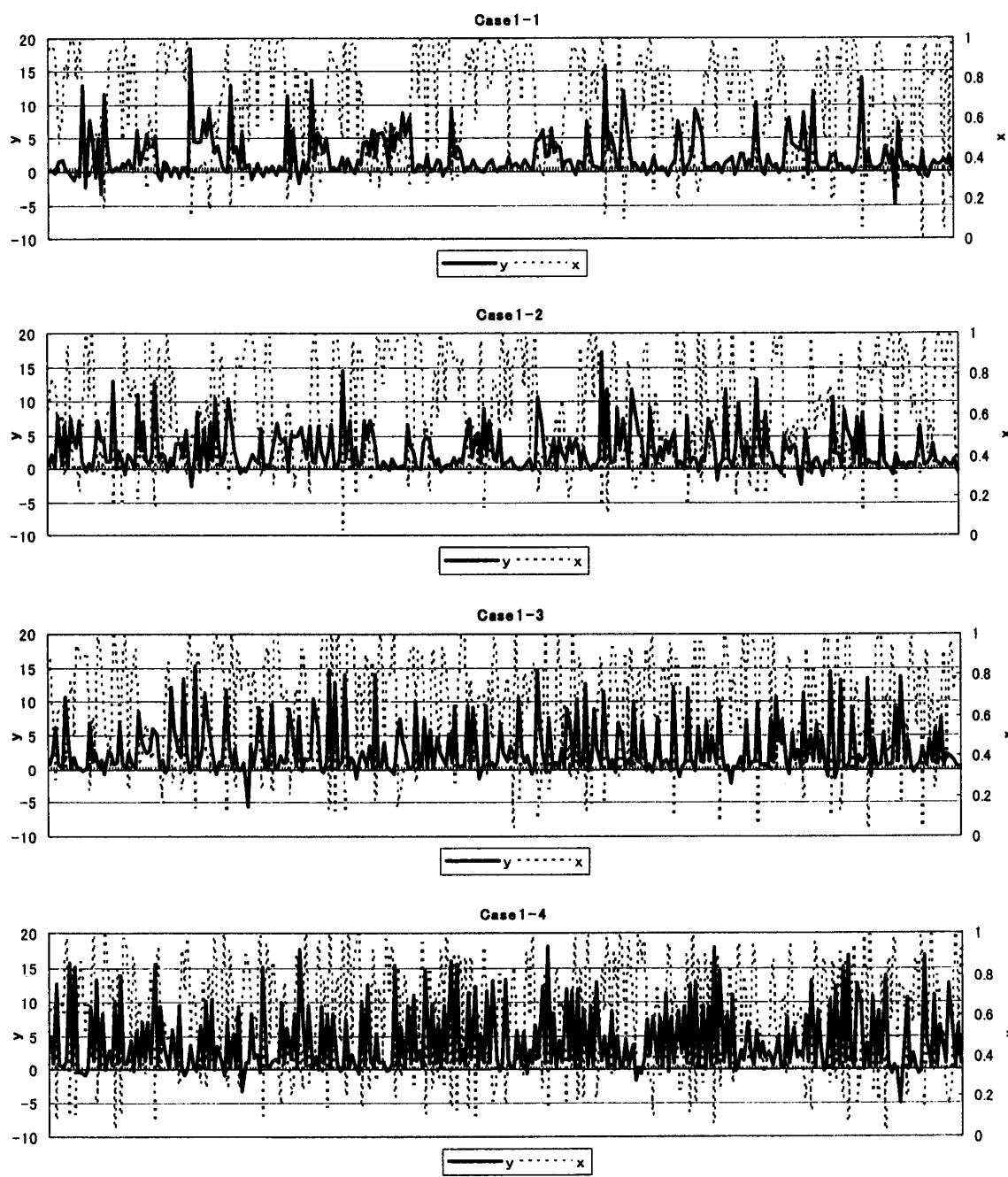


Fig. 16: Examples of Generated Time Series

Table 10: Misclassification Rate (Case1-4)

Actual Regime	AMR						out-performed rate of MSB
	MSD		MSR		MSB		
	ave.	std.	ave.	std.	ave.	std.	
Regime1	0.5304	0.1782	0.6061	0.1972	0.2995	0.1442	0.95
Regime2	0.0804	0.0455	0.0386	0.0228	0.0158	0.0123	0.94
Regime3	0.1076	0.0547	0.0350	0.0202	0.0164	0.0124	0.92
Regime4	0.6847	0.2814	0.3297	0.1398	0.2730	0.1277	0.85
Total	0.1448	0.0375	0.0791	0.0224	0.0426	0.0152	0.98

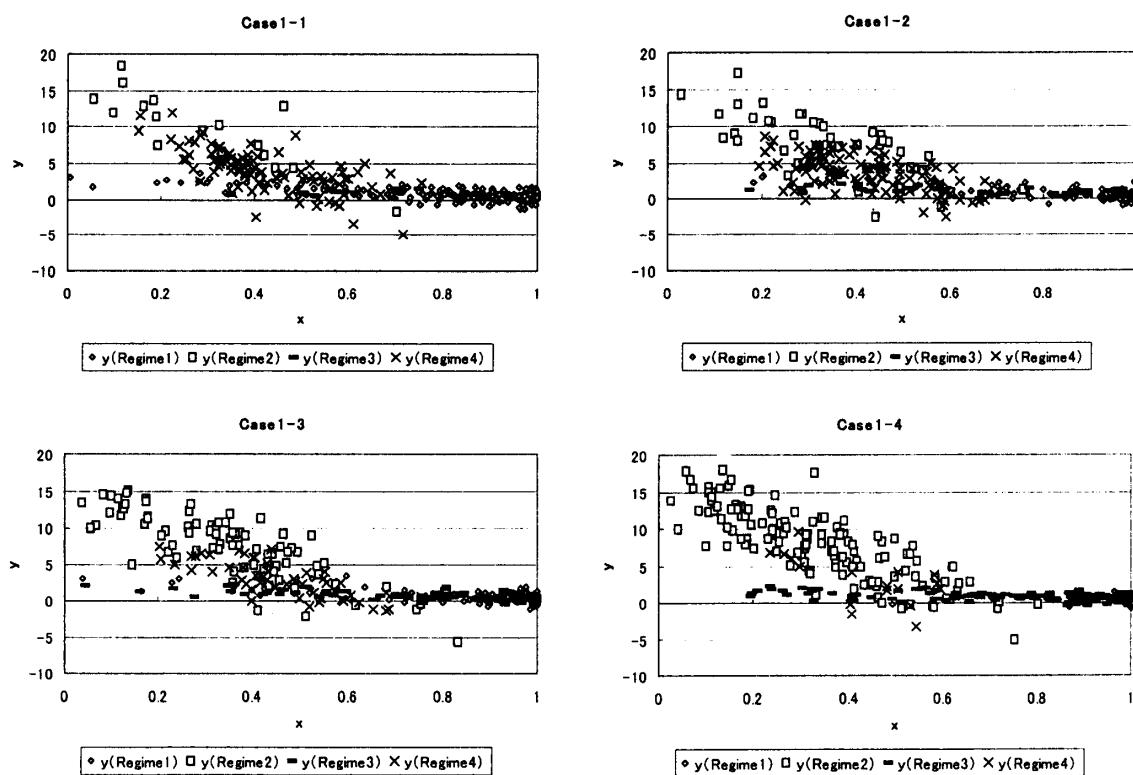


Fig. 17: Example of Scatter Plot of Generated Data

4.2 外生変数の発生機構に関する検証

本稿で提案したモデルは、変数 x が各状態毎に異なる分布に従っており、その分布から一つの x が得られるとそれにともなって y が決定されるという構造を仮定した。本小節では、変数 x の従うベータ分布を各状態毎に変化させることによって、モデルによる判別の誤分類率がどの程度変化するかについて検証した。本検証は、変数 x が変数 y に与える影響の仕方に各状態毎に差がある場合とない場合の 2 つに分けて検討している。Table 11 には、2 つの検証ケースで共通に仮定される変数 x の従うベータ分布の各状態毎のパラメータ設定を示している。表中の γ_i, δ_i ($i = 1, \dots, 4$) は(3.3)式で定義されるベータ分布の第 1 パラメータ、第 2 パラメータを示している。Fig.18 には Table 11 のパラメータ設定の場合のベータ分布を図示している。さらに、設定した外生変数の発生機構と本稿で提案したモデルによる判別の関係を評価するために、はじめに各状態毎のベータ分布の平均を(4.11)式により算定し、それを用いて(4.12)式により平均間距離の平均（Average of Distance between Means, 以降 ADM）を定義する。

(ベータ分布の平均)

$$(4.11) \quad \bar{m}_i = \frac{\gamma_i}{\gamma_i + \delta_i}.$$

(ADMの算定式)

$$(4.12) \quad ADM \equiv \frac{\sum_{j>i} |\bar{m}_j - \bar{m}_i|}{6}.$$

ここで、 \bar{m}_j は状態 j のベータ分布の平均を示し、 $|\bar{m}_j - \bar{m}_i|$ は状態 j と状態 i の平均間距離を示している。Table 12 は Table 11 のパラメータから得られる平均 \bar{m}_i の各状態毎の値を用いて算定される ADM である。ADM の数値が大きくなると相対的にみて、分布の平均の差が大きい設定になっている。

さらに、以降の検証においてはシステムモデルとして 2 つのケースで共通に下記の式を仮定した。

(仮定したシステムモデル)

$$(4.13) \quad P = \begin{pmatrix} 0.7 & 0 & 0.7 & 0 \\ 0.3 & 0 & 0.3 & 0 \\ 0 & 0.3 & 0 & 0.3 \\ 0 & 0.7 & 0 & 0.7 \end{pmatrix}.$$

4.2.1 回帰係数に差がない場合

本小節には、状態 1 と状態 3 間及び状態 2 と状態 4 間それぞれで回帰パラメータに差がないと仮定した場合の判別精度の検証結果を示す。変数 x の発生機構は前述したパラメータ設定に従うものとする。回帰パラメータの設定は Table 13 に示している。それらのパラメータを用いて、ケース毎に発生された変数の散布図を Fig.19 に示す。これから分かるように、状態 1 と状態 3 及び状態 2 と状態 4 の変数 y の散らばりは、いずれのケースにおいても同程度である。変数 x の散らばりは状態 1 と状態 4 に着目すると、ケース 2-1-3, 2-1-5, 2-1-7 における分布状況の違いが他のケースよりも大きい。また、状態 2 と状態 4 に着目するとケース 2-1-3, 2-1-8, 2-1-9 における分布状況の差が他のケースよりも大きい。

Table 14 に示すように、予想されるケース 2-1-1 を除いて全てのケースで、MSR → MSD → MSB モデルの順に誤分類率が小さくなっている。ケース 2-1-1 のケース設定

Table 11: True Parameters of Beta Distribution(Every Case)

Case	Regime1		Regime2		Regime3		Regime4	
	γ_1	δ_1	γ_2	δ_2	γ_3	δ_3	γ_4	δ_4
Case2-*1	1.79	0.46	5.64	8.05	1.79	0.46	5.64	8.05
Case2-*2	1.79	0.46	3.62	8.05	1.79	0.80	5.64	8.05
Case2-*3	1.79	0.46	2.14	8.05	1.79	1.24	5.64	8.05
Case2-*4	1.79	0.46	5.64	8.05	1.79	0.80	5.64	8.05
Case2-*5	1.79	0.46	5.64	8.05	1.79	1.24	5.64	8.05
Case2-*6	1.79	0.46	3.62	8.05	1.79	0.46	5.64	8.05
Case2-*7	1.79	0.46	3.62	8.05	1.79	1.24	5.64	8.05
Case2-*8	1.79	0.46	2.14	8.05	1.79	0.46	5.64	8.05
Case2-*9	1.79	0.46	2.14	8.05	1.79	0.80	5.64	8.05

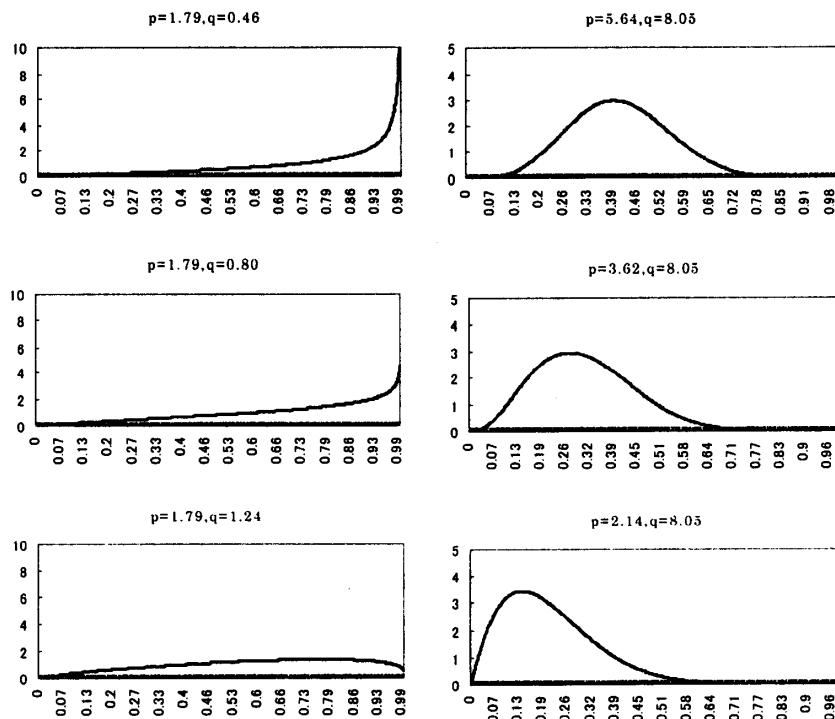


Fig. 18: Beta Distributions

Table 12: Average of Distance between Means of Distribution

Case	ADM
Case2-*1	0.2557
Case2-*2	0.2892
Case2-*3	0.3226
Case2-*4	0.2383
Case2-*5	0.2216
Case2-*6	0.3066
Case2-*7	0.2725
Case2-*8	0.3567
Case2-*9	0.3393

は、状態1と状態3間及び状態2と状態4間において変数 x の従う分布と回帰パラメータの設定ともに同じであるため、この結果も妥当であると考えられる。またOPRからも、ここで設定したケースにおいて最も判別率の高いモデルはMSBモデルである。そこから示唆されるように、MSD及びMSBモデルでは概ねADMが大きくなればAMRが小さくなる傾向にある。MSRモデルは、変数 x の分布の情報は活用していないため当然ながらその傾向は出ていない。すなわち、MSRモデルは本小節で仮定したような発生機構から得られるデータを判別するには適していない。

Table 13: True Parameters of Regression Model(Case2-1-*)

Regime	Parameter	Value
Regime1	$\sigma_{1,true}^2$ or $\sigma_{3,true}^2$	0.44
.and.	$\beta_{10,true}$ or $\beta_{30,true}$	2.74
Regime3	$\beta_{11,true}$ or $\beta_{31,true}$	-2.46
Regime2	$\sigma_{2,true}^2$ or $\sigma_{4,true}^2$	4.80
.and.	$\beta_{20,true}$ or $\beta_{40,true}$	9.31
Regime4	$\beta_{21,true}$ or $\beta_{41,true}$	-13.77

4.2.2 回帰係数に差がある場合

本小節には、4つの状態の回帰係数が全て異なると仮定した場合の判別精度の検証結果を示す。変数 x の発生機構は前節と同様に、前述のパラメータ設定のベータ分布に従うものとする。さらに、回帰係数の設定はTable 15に示す通りである。それらのパラメータを用いてケース毎に発生された系列データの散布図を書いたものをFig.20に示す。そこから分かるように、変数 y の散らばりが状態2と状態4で大きくなっている。これは前節に示したFig.19と大きく異なっている点である。変数 x の散らばりは、前節で詳述した内容と同一の傾向である。

Table 16に示すように、同一ケース内でモデル間のAMRを比較すると、ケース2-2-6及び2-2-8を除いて、MSD→MSR→MSBモデルの順にAMRが小さくなっている。ケース2-2-6においては、MSDモデルとMSRモデルの誤分類率の差がそれほど大きくなく、ほぼ同等な判別力を持っている。また、ケース2-2-8は、変数 x の従う分布の

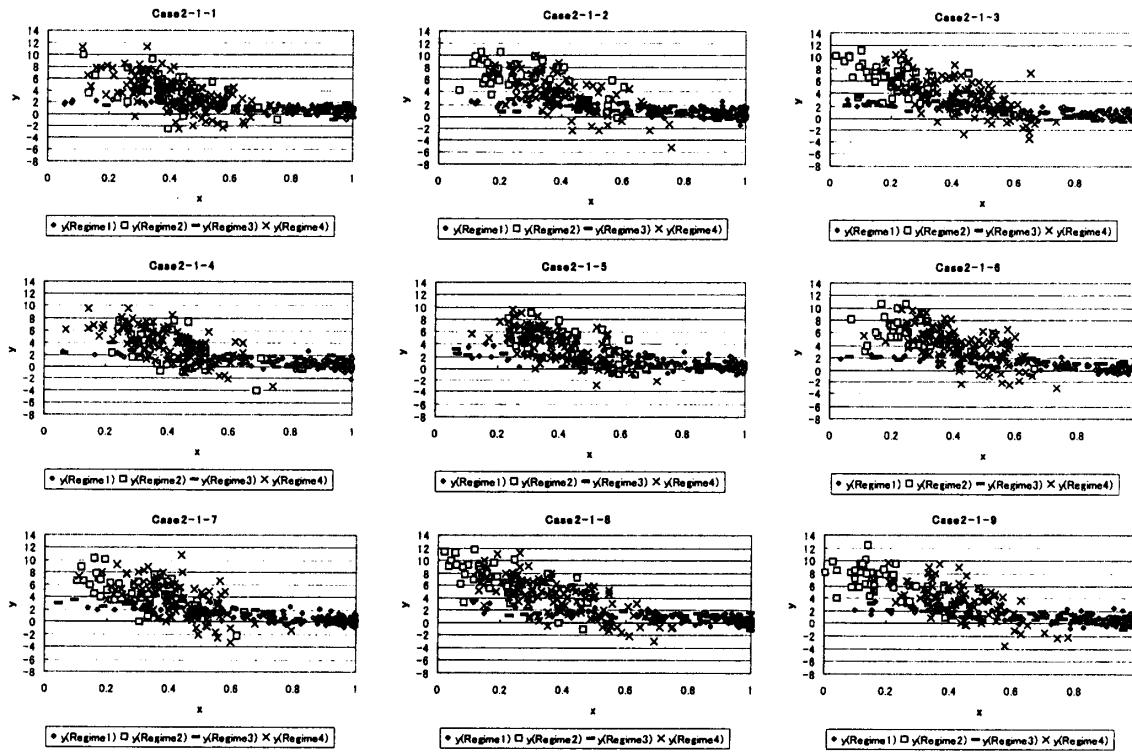


Fig. 19: Examples of Scatter Plot of Generated Data

Table 14: Misclassification Rate of Simulation2-1-*

Case	AMR						out-performed rate of MSB
	MSD		MSR		MSB		
	ave.	std.	ave.	std.	ave.	std.	
Case2-1-1	0.4177	0.0768	0.2582	0.0592	0.1601	0.0299	1.00
Case2-1-2	0.2605	0.0495	0.3936	0.0738	0.1659	0.0323	1.00
Case2-1-3	0.2506	0.0499	0.3696	0.0675	0.1508	0.0305	1.00
Case2-1-4	0.3049	0.0806	0.3922	0.0715	0.1841	0.0267	1.00
Case2-1-5	0.3382	0.0776	0.3715	0.0631	0.2048	0.0357	1.00
Case2-1-6	0.2170	0.0384	0.4227	0.0833	0.1372	0.0299	1.00
Case2-1-7	0.3035	0.0654	0.3774	0.0709	0.1758	0.0274	1.00
Case2-1-8	0.1725	0.0342	0.4219	0.0700	0.1043	0.0235	0.99
Case2-1-9	0.2219	0.0368	0.3876	0.0758	0.1340	0.0244	1.00

ADMが最も大きい設定であり、このケースでは、MSDモデルの方がMSRモデルよりも判別力の高いモデルになりうることを示唆している。全体で2ケースを除いてMSBモデルの判別力が他の2モデルよりも高く、本小節のような発生機構を持つデータに対してもMSBモデルの判別モデルとしての有意性が示された。また、設定したADMと判別力との関係をみると、MSD及びMSBモデルにおいては、概ねADMが大きくなれば誤分類率が小さくなる傾向にある。MSRモデルでは、変数 x の分布情報を使用していないためその傾向は認められない。

Table 15: True Parameters of Regression Model(Case2-2-*)

Regime	Parameter	Value
Regime1	$\sigma_{1,true}^2$	0.44
	$\beta_{10,true}$	2.74
	$\beta_{11,true}$	-2.46
Regime2	$\sigma_{2,true}^2$	4.8
	$\beta_{20,true}$	9.31
	$\beta_{21,true}$	-13.77
Regime3	$\sigma_{3,true}^2$	0.44
	$\beta_{30,true}$	2.74
	$\beta_{31,true}$	-2.46
Regime4	$\sigma_{4,true}^2$	4.8
	$\beta_{40,true}$	9.31
	$\beta_{41,true}$	-13.77

Table 16: Misclassification Rate of Simulation2-2-*

Case	AMR						out-performed rate of MSB
	MSD		MSR		MSB		
	ave.	std.	ave.	std.	ave.	std.	
Case2-2-1	0.2554	0.0546	0.2446	0.0540	0.1043	0.0228	1.00
Case2-2-2	0.2570	0.0476	0.2125	0.0366	0.1033	0.0211	1.00
Case2-2-3	0.2477	0.0343	0.1778	0.0334	0.0928	0.0210	1.00
Case2-2-4	0.2991	0.0763	0.2373	0.0467	0.1230	0.0263	1.00
Case2-2-5	0.3359	0.0768	0.2194	0.0345	0.1341	0.0257	0.99
Case2-2-6	0.2165	0.0429	0.2303	0.0443	0.0878	0.0214	1.00
Case2-2-7	0.2923	0.0506	0.1988	0.0393	0.1128	0.0253	0.99
Case2-2-8	0.1738	0.0365	0.2045	0.0480	0.0629	0.0161	1.00
Case2-2-9	0.2215	0.0444	0.1901	0.0406	0.0831	0.0256	1.00

4.2.3 数値実験のまとめ

本節では、数値実験によりマルコフ切換モデルの動的な特性及び外生変数の発生機構の違いによる判別特性の違いに関して検証した。動的な特性に関する検証からは、マルコフ遷移行列

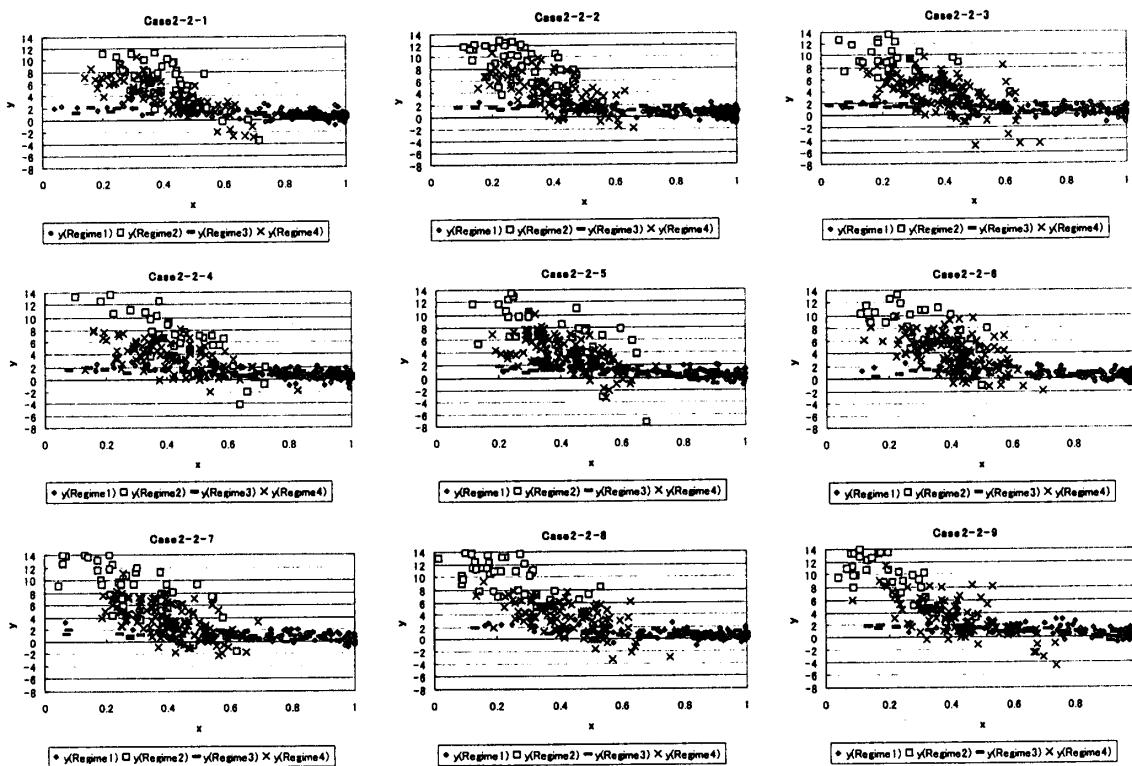


Fig. 20: Examples of Scatter Plot of Generated Data

から算定される定常確率と誤判別率の関係が示された。具体的には、特定の状態の定常確率が大きくなるような時系列データの判別では特に有用であることが分かった。また、本動的特性に関する数値実験からは、MSBモデルの判別力が他の2モデルよりも有意に高いことが示された。

外生変数の発生機構の違いによる判別率の検証においては、外生変数の発生機構を表す分布のパラメータ（今回の場合はベータ分布）を固定し、回帰係数に差がない場合とある場合それぞれにおいて検証した。本検証結果からは、前者の設定の場合、外生変数の発生分布の差が大きくなればMSD及びMSBモデルの判別力は向上することが示された。また、その設定の場合、MSRモデルによる判別は他の2モデルと比較すると判別性能が低い傾向にあった。後者の設定のようなデータに対しては、基本的にMSDモデルよりMSRモデルの判別力が高いことが分かった。さらに、MSBモデルの判別力が他の2モデルよりも有意に高いことも示されたことも記載しておく。

5 まとめと今後の課題

本稿で提案したモデルは、過去の時点における非価格プロモーションの実施状況の推定を教師無しデータから行うものである。POSデータでは販売点数、販売金額、来店客数のデータは取得できる。そのため、通常売価に比較して値引きがどの程度されていたか、すなわち価格プロモーションの有無に関してはPOSデータから簡単に情報が取得できる。一方、非価格プロモーションの有無については小売店舗毎に調査により収集しない限りはデータが入手できない。しかし、非価格プロモーションの有無により大きく売上は変動するので、非価格プロモーション実施時の構造をPOSデータで通常獲得できる販売数量及び売価などのデータからの的確に捉えることが、需要予測精度の向上につながる。また、過去の時点で非価格プロモーションの実施状況を正確に把握するためには、POSデータを用いた統計的推測法が有効である。

ロモーションを実施していたかどうかを、販売点数、売価及び来店客数のデータから判別できることは、施策の効果を踏まえてマーケティング意思決定を行うためには非常に重要なことである。そこで本稿ではPOSデータより得られる点数PIと正規化売価のみから、対象店舗で非価格プロモーションが行われていたか否かを推定する方法を提案した。

モデルでは店頭での非価格プロモーション実施の有無を潜在的な状態として捉え、この状態がマルコフ切換をしていると考えた。点数PIと正規化売価の関係をマルコフ切換モデルで定式化し、非ガウス型フィルタ・平滑化を利用して、モデルのパラメータと各時点の状態の推定を行った。

実証分析により、非価格プロモーション実施の有無を推定することが試みられた。推定の結果、実務上十分活用しうる程度の推定値を獲得することが出来た。また、教師データが存在する場合の非線形判別手法である動径基底ネットワークを用いた結果と教師データが無い場合の本稿で提案したモデルによる判別分析の結果を比較し、ほぼ同程度の誤判別率を達成できることが示された。

数値実験によって、本稿で提案したモデルの動的特性・外生変数の発生機構の違いによる、判別力の変化に関しても検証を実施し、モデルの特性に関しても意味のある知見を獲得することが出来た。

今後の課題としては、前述したように本稿で提案したモデルを、非線形性を表現できるモデルや競合商品の販売量や価格の変動の影響を考慮したモデルへ拡張することがあげられる。また、マーケティング分野でのこれらのモデルの適用範囲は広いので、他のマーケティング分野での課題へも本稿で提案したモデルを活用することも考えられる。

謝辞

この研究の一部は、文部科学省科学研究費「基盤研究（A）14208025」及び「基盤研究（B）12558023」の補助を受けている。

佐藤忠彦（正会員）〒112-0012 東京都文京区大塚3-29-1、筑波大学大学院ビジネス科学研究科、2004年総合研究大学院大学数物科学研究所統計科学専攻、博士（学術）。現在、筑波大学大学院ビジネス科学研究科・専任講師。マーケティング・サイエンス、特に販売データを用いた実証研究に従事。日本応用数理学会、日本行動計量学会、応用統計学会、日本マーケティング・サイエンス学会会員。

樋口知之（非会員）〒106-8569 東京都港区南麻布4-6-7、情報・システム研究機構統計数理研究所、1989年東京大学理学系研究科地球物理学専攻博士課程修了、理学博士。現在、情報・システム研究機構統計数理研究所・モデリング研究系・時空間モデリンググループ・教授（副所長）。統計的予測の理論と応用、特にベイズモデルを利用した情報処理法の研究に従事。日本統計学会、地球電磁気・地球惑星圈学会、アメリカ統計学会、アメリカ地球物理連合、電子情報通信学会、日本バイオインフォマティクス学会、日本マーケティング・サイエンス学会会員。

北川源四郎（非会員）〒106-8569 東京都港区南麻布4-6-7、情報・システム研究機構統計数理研究所、1974年東京大学大学院理学研究科博士課程退学、理学博士。現在、情報・システム研究機構統計数理研究所・所長。非線形・非ガウス型状態空間モデル、非定常時系列解析法、計算統計学等の研究に従事。日本統計学会、アメリカ統計学会、国際統計計算協会、日本数学会、計測自動制御協会、地震学会、日本造船学会、日本金融・証券計量・工学会、統計科学研究会、応用経済時系列研究会会員。

参考文献

- [1] Akaike, H., A new look at the statistical model identification, *IEEE Trans. Autom. Contr.*, AC-19(1974), 716-723
- [2] Bernhard, S., and Alexander, J. S., *Learning with Kernels, Support Vector Machines, Regularization, Optimization, and Beyond*, The MIT PRESS, Cambridge Massachusetts London, England, (2002)
- [3] Goldfeld, S.M., and Quandt, R.E., A markov model for switching regression, *Journal of Econometrics*, 1(1973), 3-16
- [4] Gordon, K., and Smith, A.F.M., Modeling and monitoring discontinuous change in time series, In: J.C.Spall(ed), *Bayesian Analysis of Time Series and Dynamic Linear Model*, New York Mercel Dekker, (1988), 359-392
- [5] Gordon, K., and Smith, A.F.M., Modeling and monitoring biomedical time series, *Journal of the American Statistical Association*, 85(1990), 328-337
- [6] Hamilton, J.D.A., New approach to the economic analysis of nonstationary time series and business cycle, *Econometrica*, 57(1989), 357-384
- [7] Harrison, P.J., and Stevens, C.F., A bayesian approach to short-term forecasting, *Operational Research Quarterly*, 22(1971), 341-362
- [8] Harrison, P.J., and Stevens, C.F., Bayesian forecasting (with discussion), *Journal of Royal Statistical Society, Ser.B*38(1976), 205-247
- [9] 樋口知之, 遺伝的アルゴリズムとモンテカルロフィルタ, 統計数理, 第44巻 第1号 (1996), 19-30
- [10] Holst, U., and Lindgren, G., Recursive estimation in mixture models with markov regime, *IEEE Trans. Info. Theory*, 37(1991), 1683-1690
- [11] Kim and Nelson, *State Space Model with Regime Switching*, The MIT PRESS, Cambridge Massachusetts London, England, (1999)
- [12] Kitagawa,G., Non-gaussian state-space modeling of nonstationary time series, *Journal of the American Statistical Association*, 82(1987), 1032-1063
- [13] 北川源四郎, *FORTRAN77時系列解析プログラミング*, 岩波書店, 東京, (1993)
- [14] Kitagawa,G., and Gersh, W., *Smoothness Priors Analysis of Time Series*, Lecture Notes in Statistics, No116, Springer Verlag, New-York, (1996)
- [15] 北川源四郎, 樋口知之, 知識発見と自己組織化の統計モデル, 発見化学とデータマイニング (bit別冊) , (2000), 159-168
- [16] Komaki, F., State-space modeling of time series sampled from continuous processes with pulse, *Biometrika*, 80(1993), 417-429
- [17] Leroux, B.G., and Puterman, M.L., Maximum-penalized-likelihood estimation for independent and Markov-dependent mixture model, *Biometrics*, 48(1992), 545-558

- [18] McCullagh, P., and Nelder, J.A., *Generalized Linear models*, second edition, Chapman and Hall, London, (1989)
- [19] Pena, D. and Guttman, I., Bayesian Approach to Robustifying the Kalman Filter, In:J.C.Spall(ed), *Bayesian Analysis of Time Series and Dynamic Linear Model*, New York, Mercel Dekker, 227-254, (1988)
- [20] Quandt, R.E., A new approach to estimating switching regressions, *Journal of the American Statistical Association*, 67(1972), 306-310
- [21] 坂本慶行, 石黒真木夫, 北川源四郎, 情報量統計学, 共立出版, 東京, (1983)
- [22] 佐藤義治, 統計モデルとしてのニューラルネットワーク, 統計数理, 第44巻 第1号 (1996), 85-98
- [23] Shumway, R.H. and Stoffer, D.S., Dynamic linear models with switching, *Journal of the American Statistical Association*, 86(1991), 763-769

(2003年2月24日受付)

(2005年8月19日最終稿受付)