

## 推定マーケットデータを使用した消費財系新製品の需要予測手法

宗形 聡 齋藤邦夫 樋地正浩  
(株)日立東日本ソリューションズ 研究開発部

### 概要

製品の需要予測には、指数平滑法のように時系列データに規則性を仮定し、それを過去のデータから推定する手法がよく使用される。しかし、市場投入初期の新製品は、短い時系列データしか取得できないため、その手法の適用が困難である。従来、既存の類似製品のデータから規則性を推定し、新製品の需要予測に利用する方法が用いられているが、この方法では予測精度が低いという問題が生じている。本論文では、新製品の短い時系列データ、製品寿命、累積総需要量の3つの情報をもとに予測を行う新たな手法を提案する。本手法では新製品の短い時系列データに適合し、製品寿命で累積総需要量に到達するようなモデルの要素を求め予測値を計算する。従来手法と本手法の予測精度を比較することにより、本手法が予測精度を改善することを示す。

### Propose a demand forecasting method for newly-launched consumer products

MUNAKATA Satoshi SAITO Kunio HIJI Masahiro  
Research and development department, Hitachi East Japan Solutions, Ltd.

### ABSTRACT

Demand planners usually forecast demand of products from time series data using a statistical method as exponential smoothing. The statistical method can not be applied to forecast the newly launched products because they have only short past data. When demand planners forecast the newly launched products, they utilize the method that uses regularities estimated from the past data of the similar products although they have poor forecast accuracy. This paper proposed a demand forecasting method based on three information of the newly launched product, short time series data, the planned lifetime, and total demand quantity accumulated for the lifetime. Our method computes forecast values by determining components of a forecast model that fits to short time series data of the newly launched products and reaches total demand quantity at the lifetime. We compare our method with the conventional one which uses regularities of the data of the similar products. We show that our method improves forecast accuracy.

## 1. はじめに

製品の出荷量データや販売量データなどの時系列データを統計的に分析し、需要予測を行うことは、企業が製品の発注計画、販売計画を立案する上で重要な役割を果たしている。しかし、市場に投入し

たばかりの新製品は時系列予測に十分な期間のデータを保持していない。そのため企業は次の方法により新製品の需要予測を行っている。それは、新製品の時系列データと類似する既存製品の時系列データは同じ確率分布に従うことを仮定し、類似製品の過去データから確率分布の要素を推定して新製品を予測する方法である。しかし、この方法では予測精度が低いという問題が生じている。これは、確率分布が同一の要素を持つかどうか分からないことが原因である。

その一方で、新製品の異なる予測手法として、製品の需要量が消費者の購買行動と直結していることから、消費者の購買時期が従う確率分布を使用する手法がある。この手法では、新製品に対する購買時期の確率分布を求めて需要量を計算する。購買時期の確率分布は、どの製品に対しても同一の要素で構成され、製品の時系列データが従う確率分布の違いを考慮する必要はない。しかし、購買時期の確率分布の要素を推定できるだけの、購買行動に関するデータを収集する必要がある。これは多くの企業にとって時間と費用を要するため、直接消費者の購買時期が従う確率分布を適用することは容易ではない。

本論文では、企業が既に保持している製品の時系列データだけを用いて、新製品を予測する手法を提案する。製品の累積需要量が従う確率分布は、消費者の累積購買確率が従うものと同様な形で表現されると仮定する。その際、累積購買確率に影響する要素は累積需要量に影響する要素と見なす。累積需要量が従う確率分布の要素は、製品の累積総需要量と製品寿命から推定する。ここで、累積総需要量とは、累積需要量の確率分布について時間を無限大にしたときの極限值である。製品寿命とは、累積需要量が累積総需要量に十分近くなるときの時間である。新製品の予測値は、市場投入後ひと月(4週)程度の短期時系列データに適合し、かつ製品寿命で累積総需要量に到達するように確率分布の要素を推定して算出する。その際、新製品と類似製品の製品寿命と累積総需要量は一致すると仮定し、類似製品からそれらを推定する。提案手法により新製品の予測精度が改善することを示すため、実際のデータを使用して提案手法と従来の類似製品データが従う確率分布を利用する手法の予測精度を比較する実験を行った。その結果、従来手法で予測精度が低い新製品を提案手法で予測した場合、予測精度の改善が確認できた。

第2章では、市場投入初期の新製品を対象とした従来手法が持つ課題について触れる。第3章では、課題を解決するために提案する予測手法についての詳細を述べ、第4章で実際のデータを使用した予測精度の比較実験の結果を提示する。最後に、第5章で本論文の結論を述べる。

## 2. 類似製品の時系列データを使用した新製品の需要予測

### 2.1. 時系列データ分析による需要予測の現状

製造業をはじめとする企業は、需要予測を行うために、各製品の物流拠点への出荷データや POS データなどの実売データをデータベースとして保持している。長期に渡る時系列データを統計的なアプローチによって分析し、需要予測を行う手法は既に確立されている[5]。そのような機能をソフトウェアとして実装した需要予測パッケージ製品も存在する。需要予測パッケージ単体、またはそれが持つ機能を利用して、顧客企業に需要予測システムを導入することにより、顧客は既に保持しているデータのみを使用して容易に予測できる。データベースに保存されている多くの既存製品は、需要予測システムで予測可能である。しかし、市場投入初期の新製品の場合は、時系列データが従

う確率分布を求められるほど十分なデータを取得できていないため、新製品単独の時系列予測の精度は低くなる。そこで、需要予測パッケージを使用して新製品の需要を予測する場合は、新製品に類似した既存製品の過去の時系列データを新製品の時系列データをみなし、この時系列データの確率分布を使用して新製品の需要を予測する方法が用いられる。

## 2.2. 類似製品データの規則性を利用する時系列予測が持つ課題

新製品を類似製品の時系列データを利用して予測する場合、指数平滑法がよく用いられる。指数平滑法は、時系列データにレベル・トレンド・季節性という3つの要素を外挿し、時系列の推移をこの3つの要素により説明する予測手法である。トレンドとは時系列データの大局的な増減の傾向である。季節性とは時系列データの周期的な変動である。レベルとは時系列データからトレンドや季節性を除いたものである。指数平滑法はデータの異常値に影響を受けにくく、予測精度も安定して良いので、時系列予測手法として広く用いられている。

例えば、レベル・トレンド・季節性の3つの要素により構築される1つの指数平滑法の確率モデル(Winterモデル)は以下のように定式化される[5]。

$$F(t_0 + t) = (L(t_0) + tT(t_0)) \times \hat{S}(t_0 + t).$$

$F(t_0 + t)$ は時間 $t_0$ を起点とした $t$ 時間後の予測値を表す。右辺の $L(t_0)$ は時間 $t_0$ でのレベル、 $T(t_0)$ は時間 $t_0$ でのトレンド、 $\hat{S}(t_0)$ は時間 $t_0$ での最新の季節指数を表す。右辺の3つの要素は以下の漸化式により推定される。

$$L(t_0) = \alpha(F(t_0)/S(t_0 - s)) + (1 - \alpha)(L(t_0 - 1) + T(t_0 - 1)).$$

$$T(t_0) = \beta(L(t_0) - L(t_0 - 1)) + (1 - \beta)T(t_0 - 1).$$

$$S(t_0) = \gamma(F(t_0)/L(t_0)) + (1 - \gamma)S(t_0 - s).$$

ここで、3つのパラメータ $\alpha, \beta, \gamma$ は平滑係数と呼ばれる。 $s$ は年の期間数を表す。類似製品の時系列データが従う確率分布を利用する新製品予測では、類似製品のデータから平滑係数および $L, T, \hat{S}$ を推定し、新製品の予測値を計算する。しかし、この方法では予測精度が低い新製品が多いという問題が生じる。これは、新製品の時系列データが従う確率分布が、 $L(t), T(t), \hat{S}(t)$ で記述された類似製品データの確率分布と同一になるかわからないからである。

## 3. 推定マーケットデータを使用した新製品の需要予測

本章では、市場投入後ひと月程度の短期時系列データを持つ新製品の予測手法について述べる。類似製品のデータが持つトレンドや季節性などの規則性を利用せず、かつ既存の時系列データのみを使用する予測手法を提案する。本章以下で使用するマーケットデータとは、製品寿命と累積総需要量のことを指す。

### 3.1. 消費者購買モデルと新製品予測

時系列データをほとんどあるいは全く持たない新製品の需要予測を行うために、市場にいる消費者の製品に対する購買行動を分析し、需要予測手法を構築する研究は数多く存在する[1][8]。消費

者の累積購買確率は、指数関数的な成長曲線を描くことが経験的に知られている[3]。消費者の購買確率は、企業のマーケティング活動や消費者の選好など様々な要因に影響され、その要因をパラメータとして考慮した予測モデルも研究されている[1][8]。しかし、企業がその予測モデルを使用して新製品の需要予測を行う場合、既存の時期列データ以外にも購買確率モデルのパラメータを推定するための情報を収集する必要がある。これは企業にとって多くの費用と時間を要する作業である。

本論文では、製品の需要は消費者の購買行動と同一であると考え、その結果、製品の累積需要量は Fader, Hardie らが定式化した消費者の累積購買確率モデル[3]を用いて表すことができると考える。その際、累積購買確率に影響する要素は、累積需要量が描く曲線の形状を決定するパラメータと見なす。累積総需要量を  $N$  とすると、時間  $t$  における累積需要量  $AD(t)$  は次の式で表現される。

$$AD(t) = N(1 - \exp(-\lambda t))^\gamma. \quad \dots (1)$$

ここで、 $\lambda, \gamma \in \mathbb{R}_{>0}$  は予測曲線の形状を決定するパラメータである。 $AD(t)$  は、 $t \rightarrow \infty$  のとき  $AD(t) \rightarrow N$  となる成長曲線である。

提案する予測モデルは、新製品の短期時系列データの他に、製品寿命と累積総需要量を入力とする。式(1)から  $N$  は累積総需要量である。 $\lambda$  は製品寿命から推定され、 $\gamma$  は短期時系列データから推定される。類似製品の累積時系列データに最も適合する式(1)の形の曲線を求めることで、上記のマーケットデータは類似製品から推定可能である。よって本論文で提案する予測モデルは、既存の時系列データのみで使用可能である。推定されたマーケットデータは、類似製品の時系列データが従う確率分布とは関係しない。この意味で、従来手法の持つ問題点には該当しない。

### 3.2. 推定マーケットデータを用いた予測値の算出方法

本節では、3.1.節の式(1)を使用して新製品の需要予測を行う手法について述べる。新製品の製品寿命を  $T$  とし、新製品の時間  $t$  における累積需要量を  $d_t (t = 1, 2, \dots, \tau)$  とする。提案手法の入力値である、マーケットデータ  $N$  と  $T$  および累積需要量  $d_t$  が与えられているとする。時間  $T$  で  $N$  に十分近づくように式(1)のパラメータ  $\lambda$  を計算し、新製品の累積需要量  $d_t (t = 1, 2, \dots, \tau)$  に適合するようにパラメータ  $\gamma$  を計算する。

まず  $\lambda$  を計算する。式(1)が指数関数的な成長曲線を描くように  $\gamma = 1$  とする。 $N$  は  $AD(t)$  で  $t \rightarrow \infty$  としたときの極限值であるが、製品寿命  $T$  は有限値であることを仮定する。ゆえに  $AD(T) = aN (a < 1)$  とする。この仮定から式(1)を  $\lambda$  について解くと、解  $\lambda_N$  は以下ようになる。

$$\lambda_N = -\frac{1}{T} \log \left( \frac{N - AD(T)}{N} \right).$$

次に  $\gamma$  を変数と見なし、 $d_t (t = 1, 2, \dots, \tau)$  と上で計算した  $\lambda_N$  を使用して  $\gamma$  を計算する。最小二乗法を用いて予測曲線  $AD(t)$  と  $d_t (t = 1, 2, \dots, \tau)$  との誤差が最小となるときの  $\gamma_N$  を計算すると以下のようになる。

$$\gamma_N = \frac{\sum_{i=1}^{\tau} \log(d_i/N)}{\sum_{i=1}^{\tau} (1 - \exp(-\lambda_N i))}.$$

以上の計算結果から，任意の時間 $t$ における累積需要予測値は以下の式により計算される。

$$AD(t) = N(1 - \exp(-\lambda_N t))^{\gamma_N}, \quad t = 1, 2, 3, \Lambda.$$

### 3.3. マーケットデータの推定方法

本節では，3.1.節で述べた，類似製品から新製品のマーケットデータを推定する方法について述べる．マーケットデータの推定は，類似製品の累積時系列データに式(1)を適合させることによって実行する．式(1)の曲線が類似製品の累積時系列データに最も適合するときの，累積総需要量 $N$ とパラメータ $\lambda, \gamma$ を決定すれば良い．適合度合いは曲線の決定係数で表す．類似製品の累積時系列データを $D_i (i=1, 2, \dots, \nu)$ ，その平均値を $M$ とすると，曲線の決定係数 $R$ は以下の式で与えられる．

$$R = 1 - \left( \frac{\sum_{i=1}^{\nu} (D_i - N(1 - \exp(-\lambda i))^{\gamma})^2}{\sum_{i=1}^{\nu} (D_i - M)^2} \right).$$

決定係数が1に近いほど曲線の累積データに対する適合度は高い．上式より，右辺第2項の分子が最も0に近いときの $N, \lambda, \gamma$ を求めれば，決定係数は最も1に近いことがわかる．これは累積時系列データ $D_i$ と式(1)の非線型最小二乗法に帰着する．しかしながら，解析的に $N, \lambda, \gamma$ を求めることは困難であるため，推定するパラメータを $N, \lambda$ と $\gamma$ に分割する．図1に式(1)の累積総需要量 $N$ とパラメータ $\lambda, \gamma$ を推定する手続きの簡略なフローチャートを示す．

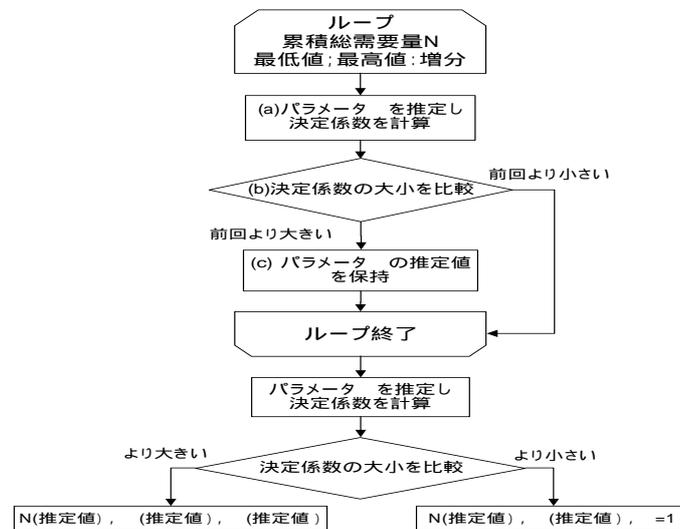


図1 マーケットデータ推定の手続きのフローチャート

まず $\gamma = 1$ に固定し， $N$ の最低値，最高値，増分を決定した後最低値から最高値まで以下の(a)，(b)，(c)の処理を繰り返して $N$ と $\lambda$ を推定する．(a)指数項を $\lambda$ の3次多項式で近似し，最小二乗法により右辺第2項の分子が最小となるときの $\lambda$ を求める．(b)曲線の決定係数を計算し，前回の処理で保持していた決定係数と大小を比較する．今回計算した決定係数の方がより1に近ければ，(c)累積総需要量 $N$ とパラメータ $\lambda$ を保持する．以上のループ処理の結果，決定係数が最も1に近い曲線式(1)の累積総需要量 $N$ とパラメータ $\lambda$ が決定する．ただし，決定係数が最も1に近い曲線は，

指定した累積総需要量  $N$  の最低値から最高値の間で最も 1 に近い曲線である。さらに増分の分だけ誤差が生じるが、 $N$  と比べて十分小さい増分を設定することでこの誤差を無視できる。

次に、推定した  $N$  と  $\lambda$  を上式の右辺に代入し、右辺第 2 項の分子が最も 0 に近いときの  $\gamma$  を推定する。これも、 $\gamma$  を変数とした最小二乗法に帰着する。計算した  $\gamma$  を用いて、式(1)と類似製品の累積データ  $D_i$  に対する決定係数を計算し、ループ処理直後の決定係数と大小を比較する。決定係数がさらに 1 に近い場合は計算した  $\gamma$  を保持し、そうでない場合は  $\gamma = 1$  のままとする。以上により、式(1)の曲線が類似製品の累積時系列データに最も適合するときの、累積総需要量  $N$  とパラメータ  $\lambda, \gamma$  を決定できる。製品寿命  $T$  は  $\varepsilon > 0$  に対して以下の式を満たすものとして計算する。

$$\left. \frac{dAD(t)}{dt} \right|_{t=T} = \varepsilon.$$

#### 4. 計算機による実験

本章では、本論文で提案した予測モデルにより予測精度が改善することを示す。そのため、指数平滑法による類似製品の規則性を利用した新製品予測と、提案手法による新製品予測の精度を比較する実験について述べる。

数千の製品出荷データからランダムに 20 品目を選択し、この 20 品目を新製品と見なして指数平滑法および提案手法で予測を行った。類似製品は、予測対象の新製品と同一シリーズの製品で、製品名称や出荷開始時期が近いものと同じ製品出荷データ群の中から選択した。新製品、類似製品ともに 104 週分の出荷量を持つ時系列データである。どちらの予測手法の場合も、新製品のデータは  $\tau = 4$  とし、出荷開始後 4 週分を使用する。5 週目から 52 週目までの 48 週分を予測し、予測期間における新製品の累積実績値と累積予測値を比較して予測精度を求める。予測精度の指標は平均絶対誤差率とする。ここで 48 週分の平均絶対誤差率は、時刻  $t$  での実績値を  $A(t)$ 、予測値を  $F(t)$  とするとき以下の式で定義する。

$$100 \times \left( \sum_t |A(t) - F(t)| / A(t) \right) / 48.$$

実験ではまず、提案した予測手法による新製品予測を行った。3.3.節に従って類似製品から新製品のマーケットデータ  $N, T$  を推定する。類似製品データは  $v = 104$  とし、出荷開始後 104 週分を使用した。累積総需要量  $N$  の最低値、最高値、増分はそれぞれ  $D_v, 10^7, 50$  とした。これは、実際のデータから多くの製品の累積総需要量は  $10^7$  以下になり、50 程度の推定誤差は無視できると仮定したことによる。ただし、最高値が  $10^7$  以上と思われる製品の場合は、十分大きな値として  $4 \cdot 10^8$  を選択した。製品寿命  $T$  は  $\varepsilon = 0.2$  として計算した。新製品の累積データ  $d_i (i = 1, 2, \dots, \tau)$  と推定マーケットデータ  $N, T$  を使用して式(1)のパラメータ  $\lambda, \gamma$  を推定し、予測値および平均絶対誤差率を計算した。

次に、指数平滑法による類似製品の規則性を利用した新製品予測を行った。新製品の 4 週分の時系列データの以前に、類似製品 104 週分の時系列データが新製品の仮想的な過去データとして存在するとし、指数平滑法を適用した。

一例として、予測対象 20 品目の中から新製品(ラベル C)の予測結果グラフを図 2 に示す。新製品 C の実績値と提案手法で算出した予測値との平均絶対誤差率は 10%であった。新製品 C の実績値と指数

平滑法による類似製品の規則性を利用した予測値との平均絶対誤差率は 32%であった．それぞれの方法で予測した 20 品目の平均絶対誤差率の結果を表 1 に示す．

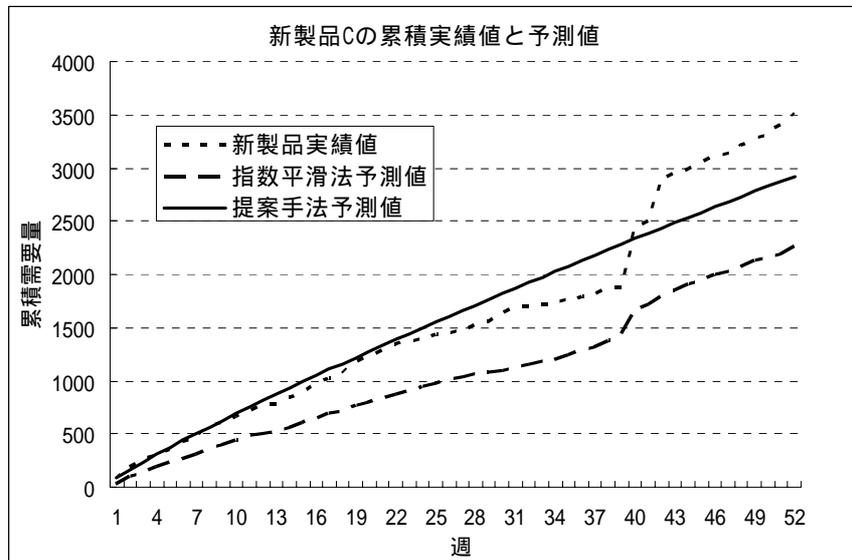


図 2 新製品 C の累積実績値と予測値

表 1 予測対象 20 品目の平均絶対誤差率の比較

項番	指数平滑法 平均絶対誤差率(%)	提案手法 平均絶対誤差率(%)	項番	指数平滑法 平均絶対誤差率(%)	提案手法 平均絶対誤差率(%)
1	333%	78%	11	32%	10%
2	58%	26%	12	18%	47%
3	98%	77%	13	50%	48%
4	73%	36%	14	34%	82%
5	65%	35%	15	80%	14%
6	324%	47%	16	52%	25%
7	47%	51%	17	126%	14%
8	47%	38%	18	82%	22%
9	10%	20%	19	99%	31%
10	17%	30%	20	12%	7%

表 1 から，半数以上の品目が平均絶対誤差率 50%を超えており，指数平滑法の意味での規則性を利用する新製品予測手法では，良い予測精度が得られる製品が少ないことがわかる．一方，提案手法で行った予測では，平均絶対誤差率 50%を超える新製品は 4 品目に減少しており，多くの新製品の予測精度が改善している．予測対象 20 品目の平均絶対誤差率の平均値は提案手法，指数平滑法それぞれ 36.9%，82.8%である．同様に平均絶対誤差率の分散は提案手法，指数平滑法それぞれ 4.9%，80.3%である．

提案手法による予測の平均絶対誤差率が指数平滑法のそれより大きな値となった品目は 5 品目ある．

この5品目(表2の項番で7, 9, 10, 12, 14)のうち3品目(9, 10, 12)については平均絶対誤差率は10~20%の範囲に入っており, 期待する予測値が得られている。しかし, 常にこのような結果が得られるように, 新製品に対する類似製品を一意に定義することは容易ではない。現状では, 製品が持つ機能や形状, 製品のターゲット, ライフサイクル, 出荷時期, 同一の製品シリーズかどうかなどの情報を参考に予測担当者が類似製品を定義している。本提案手法が持つ今後の課題は, 3品目(9, 10, 12)のように, 指数平滑法で期待する予測精度が得られた製品に対しても同等以上の予測精度になるように改善することである。本提案手法は, 新製品の市場投入後の短期実績データが得られたとき, その後の販売計画などを立案・修正する際の参考値を算出する手法として活用できる。

## 5. 結論

新製品の需要予測では, 時系列データに規則性を外挿し, 類似製品データからその規則性を推定して予測する場合, 予測精度の低い製品が多いことが問題となる。本論文では, この問題を解決するために類似製品データの規則性は使用せず, 新製品の短期時系列データ, 製品寿命, 累積総需要量を入力として予測値を計算するモデルを提案した。製品の累積需要量は, 消費者の累積購買確率モデルと同様に成長曲線を描くと仮定し, 需要予測モデルを構築した。提案した予測モデルでは, 製品寿命と累積総需要量を類似製品から推定でき, 既存の時系列データのみで予測できる。これにより, 企業は予測に必要なデータを収集する費用と時間を節約できる。実際のデータを使用した計算機実験によって, 従来手法では予測精度の低い新製品に対して, 本手法は予測精度を改善することを確認した。

## 参考文献

- [1] Gary L. Lilien, Arvind Rangaswamy, “Marketing Engineering”, Addison-Wesley Educational Publishers Inc., (1998).
- [2] Michael D. GEURTS, James E. REINMUTH, “New product sales forecasting without past sales data”, European Journal of Operational Research 4, (1980) 84-94.
- [3] Peter S. Fader, Bruce G. S. Hardie, Robert Zeithammer, “Forecasting New Product Trial in a Controlled Test Market Environment”, Journal of Forecasting, forthcoming, (2003).
- [4] Peter S. Fader, Bruce G. S. Hardie, Robert Stevens, Jim Findley, “Forecasting New Product Sales in a Controlled Test Market Environment”, <http://www-marketing.wharton.upenn.edu/ideas/pdf/Fader/Fader%20Hardie%20Stevens%20Findley%200703.PDF> (2003).
- [5] Robert L. Goodrich, “Applied Statistical Forecasting”, Business Forecast Systems, (1992).
- [6] 山田昌孝, “新製品普及モデル”, オペレーションズ・リサーチ, Vol.39, No.4, (1994) 189-195.
- [7] 棚橋菊夫, 永長亥佐夫, “新製品のプリテスト市場予測”, オペレーションズ・リサーチ, Vol.34, No.9, (1989) 477-481.
- [8] 古川一郎, 守口剛, 阿部誠, “マーケティング・サイエンス入門”, 有斐閣アルマ, (2003).