

# 大型小売店における売上予測 (第5報)

## Sales Forecast in a Large Size Retail Store (the 5th report)

橋本郁郎

Ikuro HASHIMOTO

Abstract A Sales forecasting plays important roles for making a management plan and proceeding to research development and product planing. There are various forecasts such as technical forecast, economic forecast, sales forecast and product forecast. Succeeding to previous report, the present study discusses further sales forecast of department stores in Aichi Prefecture by means of a multiple regression analysis. In the previous report, such dependent variables were used as disposable income, sales floor area and time. The present study added a dummy variable to the dependent variables for the improvement of autocorrelation of residual. As the result of the analysis, the auto correlation of residual are improved very much in which the forecast error rates are 7.86% max. and 0.42% min.

### 1. はじめに

現代社会における企業にとって、経営計画をたてる場合や研究活動をするのに、予測は非常に重要なものとなっている。第4報において、愛知県内の百貨店の総売上に付き、1979-1991年の間の検討をおこなった。今回は、その予測手順のうち、重回帰分析にダミー変数を導入して検討を進めた。

### 2. 研究方法

1979年より1990年までの12年間の月次データにより、1991年の愛知県内の百貨店の総売上高を予測するのである。バブル崩壊の影響を少しでも緩和するため、本論では、不規則変動を除くための重回帰分析において説明変数にダミー変数の

追加導入を試みた。

### 3. 回帰モデル式の構築

前報においては、モデル式の残差に自己相関が認められ、またバブル崩壊の影響があつたとはいえ予測誤差が最大で30%もあることが判明した。そこで本報では、それらを改善するため、モデル式にダミー変数を1個または2個追加導入して解析を進めることとした。以下の統計計算はSAS<sup>3)4)</sup>によって行なつた。

#### 3.1 ダミー変数(1個)の導入

##### (1) 導入方法

季節調整済の売上高を図1に示す。この図を見ると、全体を1つの回帰式で行うよりも1987年9月を境として前半と後半を別々に回帰したほうがより良いように考えられる。そこでダミー変数を1個

導入することにより、1つの回帰式でそれを実現することにした。

(2) モデル式の構築

前報のダミー変数なしの最適モデル式に、上記のダミー変数を追加し、(1)式の如き回帰式により検討を進めた。

$$Y = a_0 + a_1 X_1 + a_2 X_2 + a_3 X_3 + a_4 X_4 + a_5 X_5 + a_6 X_6 + a_7 X_7 \dots (1)$$

- Y : 総売上 (季調済み)                    URIAGE2
- X<sub>1</sub> : 可処分所得 (季調済み)            SYOTOK2
- X<sub>2</sub> : 売場面積                            MENSEKI
- X<sub>3</sub> : 時間                                 TIME
- X<sub>4</sub> : ダミー変数                         DAMY
- X<sub>5</sub> : SYOTOK2 \* DAMY                 DX1
- X<sub>6</sub> : MENSEKI \* DAMY                 DX3
- X<sub>7</sub> : TIME \* DAMY                     DX4

(1) 式の重回帰分析の結果を表1に示す。ここに [ダミー変数] \* [説明変数] の PROB > |T| の値の大きいものは、あまり導入の意味がないので、

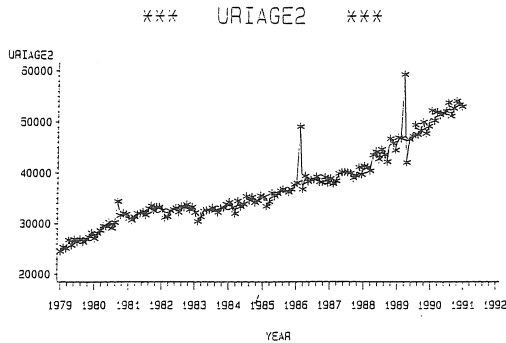


図1 季調済み売上データ

DX1, DX3を除いてこれを最適モデルとして(2)式に示す。

$$Y = b_0 + b_1 X_1 + b_2 X_2 + b_3 X_3 + b_4 X_4 + b_7 X_7 \dots (2)$$

(2) 式の重回帰分析の結果を表2に示す。

3.2 ダミー変数(2個)の導入

(1) 導入方法

前項ではダミー変数を1個導入した場合につき述べたが、ここではもう1個追加して、改めてモデル

式の構築を行うことにする。図1を見るかぎり、ダミー変数で変化させるポイントが2箇所と見られる。そこで1982年の2月と1987年の9月の2箇所で関数が変化するようにダミー変数を2個導入して解析する。

(2) モデル式の構築

前報のダミー変数なしの回帰式にダミー変数2個を追加したモデル式(3)により検討を進める。

$$Y = c_0 + c_1 X_1 + c_2 X_2 + c_3 X_3 + c_4 X_4 + c_5 X_5 + c_6 X_6 + c_7 X_7 + c_8 X_8 + c_9 X_9 + c_{10} X_{10} + c_{11} X_{11} \dots (3)$$

- X<sub>4</sub> : ダミー変数1                         DAMY1
- X<sub>5</sub> : ダミー変数2                         DAMY2
- X<sub>6</sub> : SYOTOK2 \* DAMY1                 D1X1
- X<sub>7</sub> : SYOTOK2 \* DAMY2                 D2X1
- X<sub>8</sub> : MENSEKI \* DAMY1                 D1X3
- X<sub>9</sub> : MENSEKI \* DAMY2                 D2X3
- X<sub>10</sub> : TIME \* DAMY1                    D1X4
- X<sub>11</sub> : TIME \* DAMY2                    D2X4

(3) 式の重回帰分析の結果を表3に示す。前項と同様に PROB > |T| の大きい D1X1, D1X

表1 ダミー1個の場合の重回帰分析

PARAMETER ESTIMATES						
VARIABLE	OF	PARAMETER ESTIMATE	STANDARD ERROR	T FOR H0: PARAMETER=0	PROB >  T	STANDARDIZED ESTIMATE
INTERCEP	1	-3232.924554	9487.0147598	-0.341	0.7338	0.00000000
SYOTOK2	1	-0.002609	0.00982842	-0.265	0.8067	-0.01514646
MENSEKI	1	93.242651	29.27341630	3.184	0.0018	0.40640822
TIME	1	77.466501	17.96588874	4.312	0.0001	0.42535993
DAMY	1	-3717.244813	11907.594114	-0.312	0.7556	-0.21993102
DX1	1	-0.51552	0.01489177	-0.976	0.3307	-0.36593320
DX3	1	-65.071185	37.80402753	-1.192	0.2353	-1.12205678
DX4	1	252.835920	46.13804870	5.496	0.0001	1.07465381

3, D2X3を除いて、これを最適モデルとして(4)式に示す。

$$Y = d_0 + d_1 X_1 + d_2 X_2 + d_3 X_3 + d_4 X_4 + d_5 X_5 + d_7 X_7 + d_{10} X_{10} + d_{11} X_{11} \dots (4)$$

(4) 式の重回帰分析結果を表4に示す。

4. 予測結果

1991年度の売上予測をするには、本來說明変数も予測値を用いるのであるが、本論の場合は予測式の適合度を見るため実績値を用いた。

表2 最適モデルの重回帰分析 (ダミー1個)

ANALYSIS OF VARIANCE						
SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PROB>F	
MODEL	5	7767623810.6	1553524762.1	441.896	0.0001	
ERROR	138	485151269.85	3515588.9119			
C TOTAL	143	8252775080.4				
ROOT MSE	1874.99038	R-SQUARE	0.9412			
DEP MEAN	37235.06250	ADJ R-SQ	0.9391			
C.V.	5.03555					
PARAMETER ESTIMATES						
VARIABLE	DF	PARAMETER ESTIMATE	STANDARD ERROR	T FOR HO: PARAMETER=0	PROB >  T	STANDARDIZED ESTIMATE
INTERCEP	1	7049.744428	6201.6009544	1.137	0.2576	0.00000000
SYOTOK2	1	-0.007836	0.00737527	-1.062	0.2899	-0.04926114
MENSEKI	1	67.031608	18.55877238	3.612	0.0004	0.29224626
TIME	1	97.381792	13.29181431	7.326	0.0001	0.53471258
DAMY	1	-21108	5353.7037530	-5.971	0.0001	-1.24839570
DX4	1	195.354797	30.34891342	6.437	0.0001	1.44753240
DURBIN-WATSON D	1.950					
(FOR NUMBER OF OBS.)	144					
1ST ORDER AUTOCORRELATION	0.017					

4. 1 ダミー変数1個の場合

最適モデルと考えられる(2)式のTC系列の予測値を図2に示す。

4. 2 ダミー変数2個の場合

最適モデル式(4)のTC系列の予測値を図3に示す。

5. 予測誤差

前項のTC系列の予測値に季節指数を掛けたものを売上予測値とし、(5)式より予測誤差を算出する。

$$\text{予測誤差} = \frac{\text{実績値} - \text{予測値}}{\text{予測値}} * 100 \dots (5)$$

(予測値 = TCの補外値 \* 季節調整指数)

またこれより計算した予測誤差を表5に示す

5)

6. 残差の検討

重回帰分析における残差の検討は、回帰式の正当性を見るうえで重要である。ここではダミー変数2個の場合について検討する。

6. 1 残差の時系列プロット

時間を横軸にとつて、そのときの残差の値を打点すると、残差の時間的変化を知ることが出来る。これを図4に示す。残差は0を中心にして上下に変動するが、これらについて以下検討する。

表3 ダミー2個の場合の重回帰分析

PARAMETER ESTIMATES						
VARIABLE	DF	PARAMETER ESTIMATE	STANDARD ERROR	T FOR HO: PARAMETER=0	PROB >  T	STANDARDIZED ESTIMATE
INTERCEP	1	-10611	42637.840299	-0.249	0.8038	0.00000000
SYOTOK2	1	-0.000884	0.01713291	-0.052	0.9589	-0.00355605
MENSEKI	1	107.592282	129.85758649	0.827	0.4089	0.46908381
TIME	1	185.845887	75.15425579	2.474	0.0144	1.02055885
DAMY1	1	-3111.814707	45118.865055	-0.069	0.9451	-0.20502911
DAMY2	1	3661.051737	43173.784610	0.085	0.9326	0.21660640
O1X1	1	0.006636	0.05947624	0.114	0.7541	0.16031222
O2X1	1	-0.016061	0.02011722	-0.798	0.4261	-0.40433761
O1X3	1	14.884036	138.65319854	0.107	0.9167	0.35940082
O2X3	1	-59.401016	131.79440306	-0.451	0.6529	-1.47880099
O1X4	1	-154.139148	85.45474476	-1.804	0.0736	-0.74946132
O2X4	1	144.456574	85.25801671	1.695	0.0925	1.07038873

(1) 傾向的な変化

右上がり、あるいは右下がりの傾向、周期的変化、時間と共に増大減少の傾向及び曲線的な傾向などはいずれも認められない。

(2) 符号

+符号n<sub>1</sub>が62個、-符号n<sub>2</sub>が82個となっている。この符号は(6)式で検定を行うものとする。

$$P = \frac{n-1}{2} - K \sqrt{n+1} \dots (6)$$

$$n = n_1 + n_2$$

有意水準	0.01	0.05
K	1.2879	0.9800

よつて有意水準1%では

$$P = \frac{144-1}{2} - 1.2879 \sqrt{144+1} = 55.992$$

5%では

$$P = \frac{144-1}{2} - 0.9800 \sqrt{144+1} = 59.699$$

すなわちn<sub>1</sub>、n<sub>2</sub>の小さい方の数値が、有意水準1%では55以下、5%では59以下ならばそれぞれ

表4 最適モデルの重回帰分析(ダミー2個)

ANALYSIS OF VARIANCE						
SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PROB>F	
MODEL	8	7836181903.8	979522737.98	317.421	0.0001	
ERROR	135	416593176.62	3085875.3824			
C TOTAL	143	8252775080.4				
ROOT MSE		1756.66598	R-SQUARE	0.9495		
DEF MEAN		37235.06250	ADJ R-SQ	0.9465		
C.V.		4.71777				
PARAMETER ESTIMATES						
VARIABLE	DF	PARAMETER ESTIMATE	STANDARD ERROR	T FOR H0: PARAMETER=0	PROB >  T	STANDARDIZED ESTIMATE
INTERCEP	1	2360.722807	7092.4957807	0.333	0.7398	0.00000000
SYOTOK2	1	0.005794	0.00923553	0.627	0.5315	0.03642393
MENSEKI	1	62.589961	20.12366768	3.110	0.0023	0.27288136
TIME	1	206.026365	99.70065789	4.937	0.0001	1.15126783
DAH1	1	2332.232117	1137.3364765	2.051	0.0422	0.15366441
DAH2	1	-12790	5215.3442637	-2.452	0.0155	-0.75674654
OZX1	1	-0.023363	0.01393808	-1.676	0.0960	-0.58814626
DLX4	1	-126.267372	30.12252100	-4.192	0.0001	-0.63032541
OZX4	1	105.759714	42.22934382	2.504	0.0135	0.78365423
DURBIN-WATSON D		2.227				
(FOR NUMBER OF OBS.)		144				
1ST ORDER AUTOCORRELATION		-0.114				

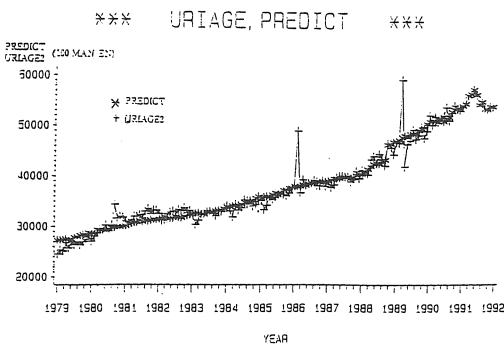


図2 予測値及び実績値(ダミー1個)

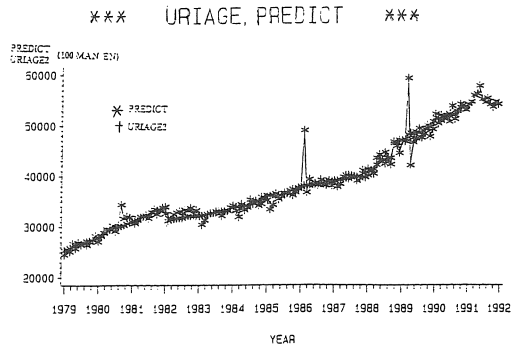


図3 予測値及び実績値(ダミー2個)

有意となり、50:50とは認められない。本論の場合には小さい方の数値は  $n_1 = 62 > 55$  となり1%の有意水準で有意ではなくなり、+-は50:50と認められる。

(3) 連の数による検定

+の連、-の連を合わせた全体の連の数を  $n_R$  で表すと、+、-がランダムにでる場合は  $n_R$  は1つの確率変数になり、ある分布に従う。

$n_1 > 10$ ,  $n_2 > 10$  の場合は連の数の分布は近似的に(7)式の正規分布に従う。

$$\mu = \frac{2n_1n_2}{n_1+n_2} + 1$$

$$\sigma^2 = \frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1+n_2)^2(n_1+n_2-1)} \quad (7)$$

従って(8)式の如く  $u$  は標準正規分布に従う。

$$u = \frac{|n_R - \mu| - 0.5}{\sigma} \dots (8)$$

これに  $n_1 = 62$ ,  $n_2 = 82$ ,  $n_R = 70$  を代入すれば、 $\mu = 71.61$ ,  $\sigma^2 = 34.37$ ,  $u = 0.185$  となり、 $u$  は標準正規分布の両側1%の点である2.576より小さくなるので有意ではない。即ち連の数から見ると、この残差系列はランダムと認められる。

(4) 連の長さによる検定

+、-の符号がランダムな系列では、一方の符号だけが連続して多く現れることはまれである。7個以上も同じ符号がならぶ場合はランダムとは見せないとする、本論の場合は+側で10個、-側で10個が連続して現れているので、連の長さからは少し問題を含んでいると考えられる。

6.2 残差の異常値

表4の分散分析表より、誤差の不偏分散  $V_e$  は  $V_e = 3085875.3824$

$$\sqrt{V_e} = 1756.6659$$

がえられるから、残差の3シグマ限界を求めると

$$\pm 3\sqrt{V_e} = \pm 5269.9979$$

表5 予測誤差

GOS1	GOS2	GOS3
11.2682	3.99585	3.95949
6.1621	-1.04990	-1.33307
6.8686	2.57814	3.00727
4.6403	0.08096	0.42199
4.6978	0.33626	-0.55135
9.7149	3.28027	5.38310
19.3700	4.93605	4.56459
16.9963	2.84258	1.81615
25.2067	4.62322	3.25142
30.9721	8.39721	7.86338
27.1328	5.10531	3.55413
21.4575	-0.53397	-1.21679

GOS1:ダミー変数なし

GOS2:ダミー変数1個

GOS3:ダミー変数2個

でこれを越えるものは

No. 86 +10984.99

No. 123 +11063.91

No. 124 -6378.65

の3個で、これらは明らかに異常値である。

### 6. 3 ダービンワトソン比

残差の連なりがランダムかどうかをコンピュータで検定するにはダービンワトソン比(D)を用いる。残差が全くランダムであれば、Dの値は2.0となり、正の相関があれば2よりも小さくなり、負の相関があれば2よりも大きくなる。本論の場合はDの値は2.227となり、負の相関が認められる場合は

$$4 - D < d_L \text{ なら有意}$$

$$4 - D > d_U \text{ なら有意でない}$$

となる。統計限界は

$$d_L = 1.48(1\%) \quad 1.61(5\%)$$

$$d_U = 1.60(1\%) \quad 1.74(5\%)$$

であり、したがっていずれの統計限界でも

$$4 - D = 1.773 > d_U$$

となり、負の自己相関は認められない。

### 6. 4 残差の各種統計量

SASのUNIVARIATE PROCEDUREを用いて、残差の各種統計量を計算したものを表6に示す。

### 7. 考察

### RESID ###

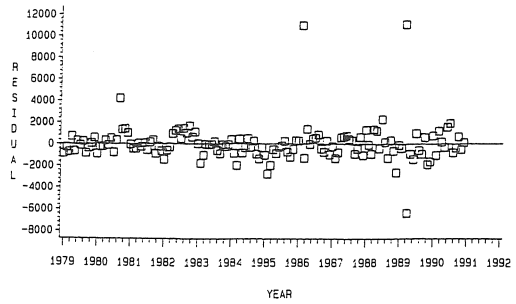


図4 残差の時系列プロット

表7 重回帰分析結果の比較

	ダミー変数なし	ダミー変数1個	ダミー変数2個	
残差	寄与率 (R <sup>2</sup> )	0.992	0.9412	0.9495
	D.W	1.500	1.950	2.227
	傾向変化	僅かに曲線傾向	曲線傾向は改善	無し
	符号	○	○	○
	連の数による検定	*	*	○
	連の長さによる検定	*	*	*

○: ランダムと認められる

\*: ランダムとは見なせず

重回帰分析において、前報におけるダミー変数なしと、本論におけるダミー変数を1個、2個導入した場合の結果を比較すると表7のごとくなる。これによると、回帰分析全体の信頼性を表す寄与率は0.9220から0.9495と向上している。また残差についても、その独立性は向上し、回帰式はより信頼を増したものと考えられる。

### 8. おわりに

企業が経営計画を立てる場合や、研究開発・製品計画等を実施するときには、予測が重要なものとなっている。百貨店の売上予測をする場合に、不規則変動を除去するための重回帰分析において、説明変数として可処分所得・売り場面積・時間をもちい、さらに残差の自己相関を改善するために、ダミー変数を1個と2個追加導入した場合につき解析をおこなった。その結果残差の自己相関は大幅に改善され、予測誤差は前報の4.64~30.97%が本論ではダミー1個の場合0.08~8.40%、ダミー

