

カオス時系列予測問題における SETAR の 性能評価とその改善

長瀬 隆久^{1,a)}

受付日 2011年11月17日, 採録日 2012年5月12日

概要: 本研究ではカオス時系列予測問題における SETAR を検証する. 従来どおりの SETAR においても, MSE などの数値としては十分な精度が実現できる. しかし, 原系列と予測系列を図示すると, いまだ改善の余地があることが明白になってくる. そこで, SETAR における区分数を増加させた局所 AR モデル, SETAR にダミー変数を加えたモデルの 2 つの手法を提案し性能評価を行った. ベンチマークとしては, 「ファーナーとシドロウィッチの方法」を簡略化した局所 AR モデルである FAR を用いた. 結論として, SETAR の区分数を増加させた局所 AR モデルにはいくつかの問題が存在し精度を向上させることが困難なケースが存在した. しかし, SETAR にダミー変数を加えると精度が改善され FAR と比べても遜色のない精度を実現することができた.

キーワード: カオス, 時系列予測, 局所リアプノフ指数, SETAR, DAR

Performance Evaluation and Improvement of the SETAR in the Problem of Forecasting Chaotic Time Series

TAKAHISA NAGASE^{1,a)}

Received: November 17, 2011, Accepted: May 12, 2012

Abstract: In this paper we examine the SETAR in the problem of forecasting chaotic time series. In the ordinary SETAR, we can realize good accuracy from looking at the MSE. But it is clear that there is room for improvement after original series and forecasting one shown by diagram. So, we evaluate the Local AR model which is increased the number of partition in the SETAR and the SETAR with dummy variables. We use the local AR model which is called the FAR, simplified the Farner and Sidorowichi's method about the benchmark. In conclusion, it is difficult improving the accuracy of the SETAR because there are some problem in the Local AR model which is increased the number of partition in the SETAR. but we show the SETAR with dummy variables can improve and get good accuracy almost like the FAR.

Keywords: chaos, time series forecasting, local lyapunov exponents, SETAR, DAR

1. はじめに

決定論に従いながらも, 初期値への鋭敏性を示し非常に複雑な振舞いをするのがカオス [1] の特徴である. 複雑なカオス時系列の予測を行う際には「ニューラルネットワーク」[2] や「ファーナーとシドロウィッチの方法」[3] など

比較的高度な専門知識を必要とするものが使用される場合が多い. しかしながら, SETAR [4], [5], [6] や DAR [7], [8] などのように単純な線形の AR モデルを応用し, 1 回もしくは数回の最小 2 乗法によりモデルを同定する簡便法タイプの手法も存在する. このような簡便法は, 瞬時に大量の品目の需要予測を行うような, きわめて実務的な場面において必要性の高いタイプの手法である. そういった状況下でカオスが発生する可能性にあらかじめ対応するためには, SETAR や DAR の研究を深めておく必要がある. しかし, これらの研究は活発に行われているとはいえ, 性

¹ 東京テレコム基礎研究所
Tokyo Telecom Basic Research Laboratory, Chuo, Tokyo
104-0061, Japan

a) d66b.v.ezjam@gmail.com

能としては未知の部分がまだ存在する．そこで，本研究ではまず，カオス時系列予測問題における SETAR の性能評価を行う．区分数が 2, 3 程度の通常の SETAR だけでなく，区分数を最大 25 にした SETAR の精度を複数のカオス時系列を対象に評価している．さらに通常の SETAR にダミー変数を加えた DAR の応用による性能改善も行った．また，ベンチマークテストとしては「ファーマーとシンドロウィッチの方法」を簡略化した FAR を用いた．本研究においては，これらの検証により SETAR の性能評価およびその改善を実現した．またそのほかにも，いくつかの有益な知見を得ることができた．

2. SETAR とその性能評価

2.1 SETAR の概要

SETAR は，「Self Exciting Threshold Auto Regressive」の略であり「しきい値 (Threshold)」により区切られた複数の区間からなる区分線形 AR モデルを意味する．推定されたモデルによりシミュレーションなどを行う想定がない場合などは「Self Exciting」による自励振動の必要がないので，単に「TAR」と呼ばれる場合もある．

今，予測を行おうとする時系列「 $X(t), X(t-1), \dots$ 」を考えたとき，区分数 = 2, 「しきい値」 = r の SETAR は以下のように表すことができる．

$$X(t) = \begin{cases} a_0 + \sum_{i=1}^{k_1} a_i X(t-i) & \text{if } X(t-d) \leq r \\ b_0 + \sum_{i=1}^{k_2} b_i X(t-i) & \text{if } X(t-d) > r \end{cases} \quad (1)$$

(a_0, b_0 は各々の区間における定数項, a_i, b_i は同様に AR 係数, k_1, k_2 は AR モデルの遅延次数, d は「しきい値」を設定する遅延次数)

また，区分数が常に 2 とは限らず，複数の「しきい値」により 3 区分以上に分割する場合もある．区分数を任意とすると j 番目の区分の AR モデルは以下のように表すことができる．

$$X(t) = a_0^{(j)} + \sum_{i=1}^{K_j} a_i^{(j)} X(t-i) \quad \text{if } r^{(j)} \leq X(t-d) < r^{(j+1)} \quad (2)$$

(K_j は j 区分目の AR モデルの遅延時数)

この集合体により予測を行うことになる．

この SETAR モデルが利用される背景には，brillinger の定理を背景とする「ラグ回帰」という概念がある（詳細については，文献 [4] の pp.176–214, 文献 [9] の pp.58–60）．これはある条件下で，現在の時系列データ $X(t)$ の条件付き期待値，すなわち $E[X(t)|X(t-i) = x]$ が x の線形関数で近似できることを保証している．このことが，任意の非

線形系を区分的に線形な関数で近似しようという理論的背景となっている．

しかしながら，実際の運用の段階では，様々な問題が存在する．「ラグ」については計算により求める方法も存在するが，膨大なデータ数を要求されたり，それほど近似精度が良くはなかったりするなどの問題がある [4], [9], [10]．また「しきい値」については，大多数の研究では 1 つ（分割領域が 2 つ）を想定しており，多くても 2 つ程度である [4], [5], [8], [10] にもかかわらず，理路整然と「しきい値」を決定するするのは容易ではない．結局，多くの場合，分析者の目視判断が何らかの形で介在している [4], [10]．

2.2 本研究における SETAR のパラメータ推定方法

SETAR におけるパラメータの推定に関して，通常 2 つの区分に分割した場合，合計 2 回の最小 2 乗法を行うことになるが，本研究では別の方式でパラメータ推定を行っている．

式 (1) の上式を第 1 区分，下式を第 2 区分とすると，式 (1) は「ダミー変数： D 」を用いて以下のように表すことが可能である．

$$X(t) = a_0 + (b_0 - a_0)D + \sum_{i=1}^K \{a_i + (b_i - a_i)D\}X(t-i) \quad (3)$$

$$D = \begin{cases} 0 & \text{第 1 区分の場合} \\ 1 & \text{第 2 区分の場合} \end{cases}$$

このように $D = 0$ のときに式 (1) の上式が， $D = 1$ のときに式 (1) の下式が出現する式を表すことができる．こういった表現は計量経済学などでは「係数ダミー」と呼ばれ，定数項ではなく係数をダミー変数で調整する分析に使用されている [11], [12]．このように SETAR は「係数ダミー」を用いると 1 つの式で表すことができる．

また， $b_0 - a_0 = a'_0$, $b_i - a_i = a'_i$ とおくと式 (3) は

$$X(t) = a_0 + a'_0 D + \sum_{i=1}^K a_i X(t-i) + \sum_{i=1}^K a'_i D X(t-i) \\ = a_0 + a'_0 D + \sum_{i=1}^K a_i X(t-i) + \sum_{i=1}^K a'_i Y(t-i) \quad (4)$$

のようにも変形することができる．さらにダミー変数が増えれば

$$X(t) = a_0 + a'_0 D + a''_0 D_2 + \sum_{i=1}^K a_i X(t-i) \\ + \sum_{i=1}^K a'_i D X(t-i) + \sum_{i=1}^K a''_i D_2 X(t-i)$$

$$= a_0 + a'_0 D + a''_0 D_2 + \sum_{i=1}^K a_i X(t-i) + \sum_{i=1}^K a'_i Y(t-i) + \sum_{i=1}^K a''_i Z(t-i) \quad (5)$$

のようにダミー変数の数に応じて項が増加するだけであり、明らかに最小2乗法は1度で済む。

ところが、上記の $Y(t-i)$, $Z(t-i)$ などは「ダミー変数 $\times X(t-i)$ 」という構造であり、ダミー変数が0のときは、 $Y(t-i)$ や $Z(t-i)$ そのものが0になるため、区分数の増加すなわちダミー変数の増加にともない説明変数の値に0が多発することになる。それにより最小2乗推定を行う際に行列演算が非常に不安定になる。

一般的に最小2乗法は、説明変数行列を X 、従属変数のベクトルを y とすると推定されるパラメータベクトル β は以下のように表される [13], [14]。

$$\beta = (X^t X)^{-1} X^t y \quad (6)$$

この説明変数行列 X に「0 (ゼロ)」が多発し、上記の式 (6) の $(X^t X)$ が特異になると、逆行列の推定、すなわち最小2乗推定が不安定になることが知られている。

この種の問題に関して計量経済学では「リッジ回帰」[15], [16], [17] という手法が存在するが非常に恣意性の強い手法である。そのため、本研究では工学分野でこの種の問題の解決法として用いられているムーアペンローズ型一般逆行列 [18], [19], [20], [21], [22], [23], [24] を採用する。

本研究における SETAR は、複数の区分を複数のダミー変数により1本の式にまとめ、そのパラメータを推定する際には逆行列の $(X^t X)^{-1}$ をムーアペンローズ型一般逆行列に置き換えて推定を行っている。この方式により区分数を増加させても演算が安定し、かつパラメータを求めるための演算は1度で済むようになる。また、実際の行列の演算に関しては MATLAB を用いて実行している。

2.3 SETAR の性能評価

本研究で性能評価の対象とするカオス時系列は文献 [5] において用いられているものに「テント写像」を加えた以下の①～⑦である。

①ロジスティック写像 1

$$X(t) = \alpha X(t-1)(1 - X(t-1)) \quad (7)$$

$(\alpha = 3.7, X(0) = 0.1)$

②ロジスティック写像 2

$$X(t) = \alpha X(t-1)(1 - X(t-1)) \quad (8)$$

$(\alpha = 4.0, X(0) = 0.1)$

③エノン写像

$$\begin{aligned} X(t) &= 1 + Y(t) - 1.4X^2(t-1) \\ Y(t) &= 0.3X(t-1) \quad (X(0) = 0.1, Y(0) = 0.1) \end{aligned} \quad (9)$$

④3次の非線形写像 [25]

$$\begin{aligned} X(t) &= 1.9X(t-1) - X^3(t-1) + Y(t-1) \\ Y(t) &= 0.5X(t-1) \quad (X(0) = 0.1, Y(0) = 0.1) \end{aligned} \quad (10)$$

⑤池田写像 [26]

$$\begin{aligned} X(t) &= q + b(X(t-1) \cos \theta - Y(t-1) \sin \theta) \\ Y(t) &= b (X(t-1) \sin \theta - Y(t-1) \cos \theta) \\ \theta &= \gamma - a/(1 + X^2(t-1) + Y^2(t-1)) \quad (11) \\ (X(0) &= 0.5, Y(0) = 0.7, q = 1.0, \gamma = 0.4, \\ a &= 6.0, b = 0.7) \end{aligned}$$

⑥指数型 AR モデル [27]

$$\begin{aligned} X(t) &= \{0.36 - 40 \exp(-X^2(t-1))\} X(t-1) \\ &\quad - \{0.99 - 0.36 \exp(-X^2(t-1))\} X(t-2) \end{aligned} \quad (12)$$

⑦テント写像

$$X(t) = \begin{cases} 2X(t-1) + \varepsilon & (0 \leq X(t-1) \leq 0.5 \text{ のとき}) \\ 2 - 2X(t-1) + \varepsilon & (0.5 < X(t-1) \leq 1 \text{ のとき}) \\ \varepsilon = 2^{-52} \end{cases} \quad (13)$$

$X(0) = 0.123$

(テント写像はビットシフト写像の一種で、そのまま退化する場合がありますので ε という微小な倍精度の計算機イプシロンを付加している)

上記の7つのカオスの生成モデルにおいて、時系列の遅延次数すなわち $X(t-n)$ の n の最大が2であるので、本研究で性能評価を行う SETAR における遅延次数はすべて2とする。また SETAR は、パラメータを推定する前に、区分化を行う「しきい値」を決める必要がある。この「しきい値」の設定は SETAR において分析者の主観が入ってしまうところである。従来法でも多くの場合、分析者の目視判断によっている。

理論的にいえば、

(i) 極大/極小点

$$\frac{dx(t)}{dx(t-n)} = 0 \quad \text{となる点 (n はラグ)}$$

(ii) 不連続点

$$\frac{dx(t)}{dx(t-n)} \quad \text{が定義できない点}$$

を「しきい値」にするのが良いはずである。(i) に関しては微分を差分で近似することによって何らかの指標を作成し、それにより判断するということが考えられる。また、(ii) に関しては⑦のテント写像などのように接している不

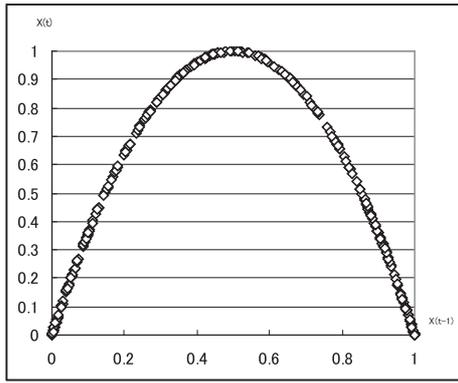


図 1 ロジスティック写像のアトラクタ
Fig. 1 The attractor of the Logistic map.

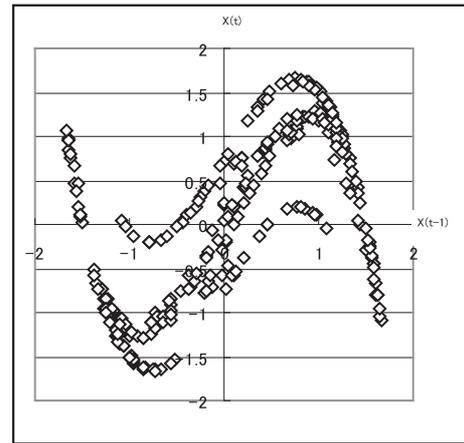


図 3 3 次の非線形写像のアトラクタ
Fig. 3 The attractor of the third-order nonlinear map.

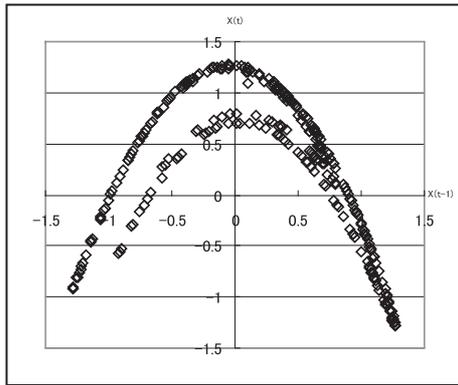


図 2 エノン写像のアトラクタ
Fig. 2 The attractor of the henon map.

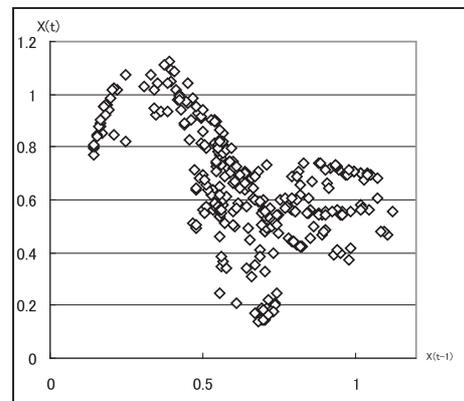


図 4 池田写像のアトラクタ
Fig. 4 The attractor of the ikeda mpp.

連続点は、左極限および右極限が同一になるので厳密には不連続点ではない。しかし区分線形の境界となる 1 点だけでも欠けていれば、厳密にはそこが不連続点となるが、離散時間では事実上 (i) と同様に扱って差し支えないであろう。もし接していない不連続点 (跳躍不動点や真性不動点など) がある場合は、離散化したデータにおいて「異常値の検出法」などを応用して判定することが考えられる。本研究では (i), (ii) の双方においていくつかの指標を作成し試行錯誤を試みたものの微分を差分で近似した誤差などのため、良好な結果は得られなかった。よって、まずは従来法どおり散布図などから目視で判断をする。

文献 [5], [8], [10] などで用いられている方法と同様に時系列の「縦軸: $X(t)$, 横軸: $X(t-1)$ 」としてプロットした図 1, 図 2, 図 3, 図 4, 図 5, 図 6, 図 7 のアトラクタの形状から判断する。前述の (i) 極大/極小点, (ii) 不連続点を目視で判断することになる。

ロジスティック写像, エノン写像のアトラクタは各々, 図 1, 図 2 であるが, これらは上に凸の 2 次関数の形状を中央の $X(t-1) = 0.5$ で区分化する。また, 3 次の非線形写像は図 3 であるが, $X(t-1) = \pm 1$ で区分化する。池田写像のアトラクタは図 4 であるが, これは $X(t-1) = 0.3, 0.7$ で区分化する。また, テント写像においては, $X(t-1) = 0.5$

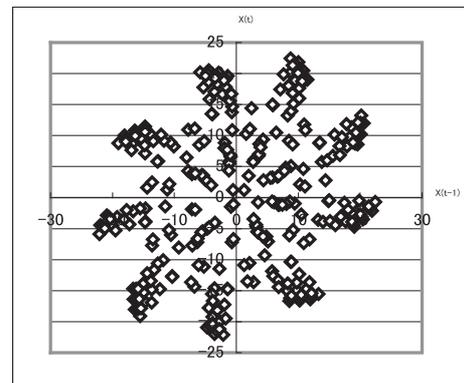


図 5 指数型 AR モデルのアトラクタ
Fig. 5 The attractor of the Exp-AR model.

で区分化する。これらに関しては、「しきい値」のラグはすべて 1 で分析可能と判断したため、ラグ 2 以降の散布図は分析を行っていない。

上記とは異なり指数型 AR モデルの場合は、図 5 のようにカオスアトラクタは存在するが、区分化可能な領域が見出せない。これは $X(t)$ と $X(t-2)$ などをプロットしても図 6 のように単純な線形性は現れるが区分化すべき箇所がなく、文献 [5] においても SETAR では推定不能として

表 1 各種 SETAR の精度

Table 1 The accuracy of various kinds of the SETAR.

MSE	AR(2)	SETAR	提案法①				提案法②
			SETAR(5)	SETAR(10)	SETAR(20)	SETAR(25)	SETAR+ダミー
ロジ3.7	0.019544146	0.005604675	0.00250807	0.000912653	0.000140408	5.99642E-05	0.000394258
ロジ4.0	0.119842505	0.063991291	0.015766026	0.005202142	0.001022446	0.000369687	0.003326705
エノン写像	0.457148319	0.028696593	0.03566549	0.004366584	0.00137297	0.006597629	0.008036747
3次写像	0.492800917	0.228323993	0.679032653	0.679032653	0.679032653	0.679032654	0.016760837
池田写像	0.024801021	0.019426284	0.01483264	0.011671617	0.009604124	0.007949406	0.017227937
指数AR	13.94971743	13.94971743	11.8838128	17.79190464	3.851447809	9.410242441	0.560004558
テント写像	0.131552708	9.19696E-29	0.001492829	5.39144E-18	5.66846E-16	0.001452556	-
ロジ3.7	0.017996509	0.007159807	0.000380195	0.000766168	0.000165223	7.21041E-05	0.000686208
ロジ4.0	0.126193503	0.428587865	0.017949238	0.005138375	0.001274306	0.000553702	0.003425615
エノン写像	0.458453531	0.026559725	0.034864451	0.004818179	0.00150383	0.006123573	0.008258009
3次写像	0.518250148	0.282886286	0.689373259	0.689373259	0.689373259	0.689373259	0.01528608
池田写像	0.021940284	0.018474858	0.014021044	0.011124661	0.009582941	0.01025088	0.017004437
指数AR	12.21092091	12.21092091	12.63572372	27.60632936	7.914605369	19.79223672	2.059949942
テント写像	0.130309036	4.57242E-29	0.001498544	1.58748E-15	5.90489E-16	0.001453486	-

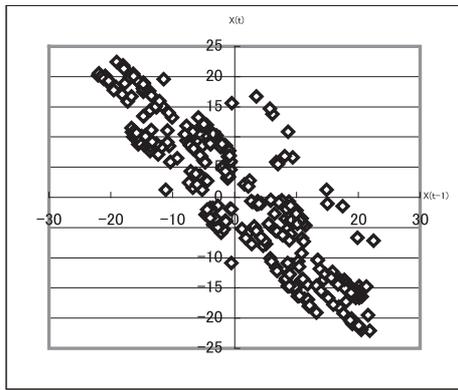


図 6 指数型 AR モデルのアトラクタ ($X(t), X(t-2)$)

Fig. 6 The attractor of the Exp-ARmodel. ($X(t), X(t-2)$).

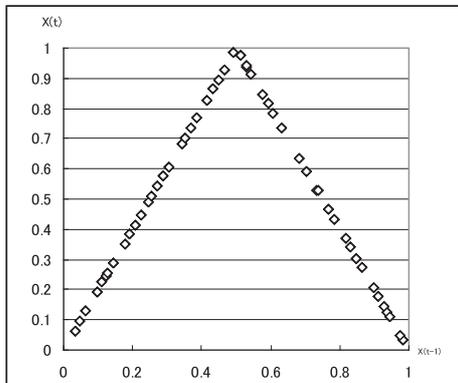


図 7 テント写像のアトラクタ

Fig. 7 The attractor of the tento map.

いる。本研究では、指数型 AR モデルには区分化しない通常の AR モデルをあてはめる。これは以下のように式 (2) の $r^{(j)}$ や $r^{(j+1)}$ を ∞ とすれば AR モデルも SETAR モデルに包含されるという考え方である。

$$X(t) = a_0 + \sum_{i=1}^K a_i X(t-i) \quad (14)$$

if $-\infty \leq X(t-d) < \infty$
(K は AR モデルの遅延時数)

以上の設定により、SETAR の推定を行った際の MSE の

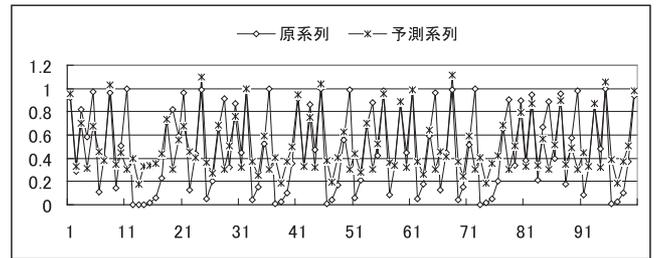


図 8 ロジスティック写像 2 における原系列と予測系列

Fig. 8 The original series and forecasting series of the Logistic map 2.

値が表 1 のデータ中の左から 2 列目である。一番左側の列は精度を比較するための簡単なベンチマークとして線形 AR モデルを用いた結果である。上段が、パラメータ推定にデータ数：300 (以下、BasicDataSet と呼ぶ) を用いた結果であり、下段は上段で推定されたパラメータを用いて新たなデータ：300 (以下、NewDataSet と呼ぶ) に対して 1 期先予測を行い、オーバーフィッティング (過剰なあてはめ) の傾向の有無の確認したものである。この MSE を見る限り、SETAR の精度は非常に良好であり、一見、十分な結果だと思われる。

ところが、図 8 は式 (9) の「ロジスティック写像 2」における原系列と SETAR による予測系列を図示したもののだが、明らかに改善の余地があることが分かる。このようにカオスは複雑な挙動を示すがノイズとは異なるため、数値的に精度が良いにもかかわらず、実際上はとらえうる挙動を逃しているという状態が存在しうる。

2.4 区分数を増加させた SETAR の性能評価

前述の結果をふまえ、SETAR の性能改善を試みる。図 8 のように予測が粗くなるのは、図 1 のような上に凸の 2 次関数などを 2 つの領域に区分化し、線形化してしまっ際に、極大点付近の緩やかなカーブと極大点から離れた急峻な直線部分を平均化した直線をあてはめていることが原因であり、平均化された推定を行うと極大付近でもそれ以外

表 2 区分化による相関の減少 (3 次の非線形写像)

Table 2 The decrease of correlation by partition (The third order nonlinear map).

SETAR	区間	$X(t-1) < -1$				$-1 \leq X(t-1) < 1$				$1 \leq X(t-1)$	
	相関係数	-0.948466472				0.877096748				-0.90902791	
SETAR(10)	区間	$X(t-1) < -1.6$	$-1.6 \leq X(t-1) < -1.2$	$-1.2 < X(t-1) \leq -0.8$	$-0.8 < X(t-1) \leq -0.4$	$-0.4 < X(t-1) \leq 0$	$0 \leq X(t-1) < 0.4$	$0.4 \leq X(t-1) < 0.8$	$0.8 \leq X(t-1) < 1.2$	$1.2 \leq X(t-1) < 1.6$	$1.6 \leq X(t-1)$
	相関係数	-0.966	-0.985	-0.295	0.248	0.300	0.531	0.264	-0.083	-0.978	-0.935

でも誤差が発生することになる。そこで、極大付近とそれ以外の推定を分割することを考えると、区分数を増加させることが自然である。本来ならば、先の「しきい値」設定の判断基準

- (i) 極大/極小点
- (ii) 不連続点

に加えて、(iii) 変曲点などを加えることが考えられるが、変曲点を求めるのは2次導関数であるため、単純に差分化すると近似が粗すぎて明らかに無理がある。かといって図から変曲点に該当する箇所を目視で選択するのは、(i)、(ii) の判断以上に危険である。

そこで本研究では、等間隔の細かい T 個の小領域に区分化する SETAR(T) を考える。たとえば、モデルによらず区分数を $T = 10$ とすれば、各々の区分内で最適な AR 係数が推定され、局所 AR モデルの集合体により予測を行うことが可能である。細かく分割すれば、事実上、極大付近とそれ以外の推定を分けることにもなり精度を向上させる可能性があるとともに、アトラクタの形状から直感的な判断で「しきい値」を選択することに比べて、恣意性を排除できるという点でも望ましい。

そこで、区分数 T は、 $T = 5, 10, 20, 25$ の4種類を用いて SETAR(T) の精度の比較検証を行うことにする。この方法を本研究における「提案法①」とする。「しきい値」については、従来法により推定した SETAR の「しきい値」を残してしまうと、区分数 T を5, 10, ... と増やした際に区間を等間隔に保つことが難しく (特に従来法の SETAR の「しきい値」が2つの場合)、近似の粗い箇所と細かい箇所が発生してしまうので、当初の「しきい」値を使うことはあまり望ましくない。そこで、提案法①の「しきい値」の設定は以下のようにする。最も両端を時系列の最大値、最小値の小数点第2位を切り捨てた値とし、たとえば最大値 = 1.42, 最小値 = -1.23 であれば、「最大のしきい値」= 1.4, 「最小のしきい値」= -1.2 とし、最大値と最小値の間を均等に区分数 T に応じて分割する。なお、区分数 5, 10, 20 は、 $X(t-1)$ の値を均等に、区分数 25 は $X(t-1), X(t-2)$ の値を各々区分数 5 として均等に区分化した。それらの「SETAR(T)」による予測精度の結果が表 1 の「提案法①」の部分である。やはり、区分数を増加させた提案法①の方がおおむね精度が向上していることが確認できる。

しかしながら、2点見逃せない点がある。1点は「指数

型 AR モデル」において、BasicDataSet では良好な精度を示す場合でも、NewDataSet においては誤差が大きくなる傾向があり、オーバフィッティングの兆候を示している。SETAR(T) の次数 T により多少の違いがあるが、全般的に指数型 AR モデルに対して過剰なあてはめが確認できる。

もう1点は、「3 次の非線形写像」において、「SETAR(T)」が線形 AR モデルよりも精度が劣っているという点である。これは意外な事実であり、非線形性に対応するために細かく区分線形化しているのだから、通常の線形 AR モデルよりも良好な精度が期待される。ところが、データ上の事実としてはそうはならない。このことの原因は以下による。

図 3 の「3 次の非線形写像」のアトラクタを見ると、 $|X(t-1)| > 1$ の区間では1本の直線もしくは曲線で近似できるような形状だが、 $|X(t-1)| < 1$ の区間では線形モデルにノイズが入ったような形状をしている。この後者の区間は、全体としてみると線形性が強いのだが、細かく区分化してしまうと各区間のデータの線形相関が落ちてしまう。参考までに実際に各区間の相関係数を計算したのが表 2 である。上段に SETAR の各区間の相関係数、下段に SETAR(10) のそれを示しているが、区分数を増加させた SETAR(10) では、各区間の $X(t)$ と $X(t-1)$ の相関係数の値が特に $|X(t-1)| < 1$ の範囲で小さくなっていることが確認できる。

本研究における相関係数とは、最も一般的なピアソンの積率相関係数である。これは帰無仮説を無相関とする場合、以下の式で t 検定を行うことにより、有意な値であるか否かを判定可能であることが知られている [28], [29]。

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \quad (15)$$

r : 標本相関係数, n : 標本数

上記の t 値に関して自由度 $n-2$ で検定することが可能であり、両側 5%水準で有意性を判断した。表中の「太字かつ斜字体」にした相関係数は、5%の有意水準で「0」と判定されたことを示している。

3 次の非線形写像の場合、従来法の SETAR を推定した際の中央の領域は、相関係数が「0.877」であったが、SETAR(10) の推定時には、0.2 や 0.3 の値が計測されても事実上は、ほとんど無相関の状態になってしまっている。ここに結果は示さないが、その他の SETAR(T) に関してほぼ同様である。このような場合、区分数を増加させた SETAR よりも単なる線形 AR モデルの方が精度が良いと

表 3 区分化による相関の減少 (池田写像)

Table 3 The decrease of correlation by partition (The Ikeda map).

SETAR	区間	$X(t-1) < 0.3$			$0.3 \leq X(t-1) < 0.7$				$0.7 \leq X(t-1)$		
	相関係数	0.654410951			-0.728651584				0.510744176		
SETAR(10)	区間	$X(t-1) < 0.1$	$0.1 \leq X(t-1) < 0.2$	$0.2 \leq X(t-1) < 0.3$	$0.3 \leq X(t-1) < 0.4$	$0.4 \leq X(t-1) < 0.5$	$0.5 \leq X(t-1) < 0.6$	$0.6 \leq X(t-1) < 0.7$	$0.7 \leq X(t-1) < 0.8$	$0.8 \leq X(t-1) < 0.9$	$0.9 \leq X(t-1)$
	相関係数	-	0.955	-0.122	0.310	-0.550	-0.104	-0.464	0.465	0.041	-0.031

表 4 区分化による相関の減少 (エノン写像)

Table 4 The decrease of correlation by partition (The Henon map).

SETAR	区間	$X(t-1) < 0$					$0 \leq X(t-1)$				
	相関係数	0.887155194					-0.950458218				
SETAR(10)	区間	$X(t-1) < -1.2$	$-1.2 \leq X(t-1) < -0.9$	$-0.9 \leq X(t-1) < -0.6$	$-0.6 \leq X(t-1) < -0.3$	$-0.3 \leq X(t-1) < 0$	$0 \leq X(t-1) < 0.3$	$0.3 \leq X(t-1) < 0.6$	$0.6 \leq X(t-1) < 0.9$	$0.9 \leq X(t-1) < 1.2$	$1.2 \leq X(t-1)$
	相関係数	1.000	0.379	0.477	0.555	0.252	-0.204	-0.638	-0.824	-0.923	-0.949

いう現象が確認された。

図 3 の「3 次の非線形写像」で上記のような現象が起きるならば、図 4 の「池田写像」でも同様の現象が起きるのではないかと考えられる。特に池田写像は 3 次の非線形写像にデータの散らばり具合が類似しているのも最も懸念される。そこで表 3 に池田写像における相関係数の一覧を同様に示した。表 2 と同様に相関係数の t 検定も行ってある。表 2, 3 を見る限り全体の中では、10 区間中 5 区間において相関係数が「0」であろうと判断された点については 3 次の非線形写像も池田写像も同じである。しかしながら、2 つの違いがみられる。1 点目は 3 次の非線形写像においては、従来法の SETAR の際の相関がすべての区間において「0.9」程度でかなり高い値を示している状態から、5 区間が無相関の状態になったことが分かるが、表 3 の池田写像の場合は従来法の際の相関がそこまで高い値を示していない。つまり相関の減少の落差が異なるというわけである。2 点目は、池田写像の場合「 $0.1 \leq X(t-1) < 0.2$ 」の区間においてかえって相関が上昇していることである。このような状況下において池田写像では SETAR(T) による推定において精度が劣化することがなかったといえる。また、参考までに表 4 にエノン写像における同様の分析を示したが、SETAR において相関が高く、SETAR(10) においても無相関と推定される領域は少なかった。

区分数を増加させた「SETAR(T)」(提案法①)の性能評価を行った結果をまとめると以下のようになる。「しきい値」の値を分析者の主観によらずに機械的に決定でき、かつよりきめ細かく領域を分割して精度向上を目指しても、一般逆行列を用いた推定で安定的な運用が見込めた「SETAR(T)」であったが、区間の細分化により線形相関がかえって崩れる現象が存在していることが確認されたため、運用には慎重を期さなければならないことが判明した。

3. SETAR にダミー変数を加えたカオス時系列予測法

「提案法①」は容易に自動化が可能であり、分析者の恣意性が入りやすく、写像の非線形性もとらえうるという期

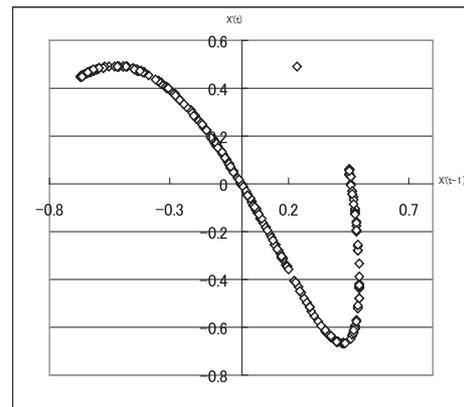


図 9 ロジスティック写像の 1 次階差

Fig. 9 The first-order difference of the Logistic map.

待があったものの、汎用性の面で思わぬ盲点があった。そこで、「提案法②」として SETAR にダミー変数を加えた予測法の検証を行う。

3.1 AR モデルにダミー変数を加えたカオス時系列予測法

カオス時系列予測に AR モデルを用いた簡便法としては SETAR のほかに DAR が知られている [7], [8]。DAR とはカオス時系列に何らかの形で線形性を見出し、それ以外の部分にダミー変数を設定するというシンプルな手法である。たとえば、「①ロジスティック写像 1」の 1 次の階差系列を $X'(t)$ とし、 $X'(t)$ と $X'(t-1)$ の相関図を描くと図 9 のように、原点付近に原系列よりも線形性が明白に出現することが知られており [7]、この線形部分に AR モデルをあてはめそれ以外の部分をダミー変数で処理する [7]。また、さらに複雑なケースに対応するためには 1 次だけでなく 2 次などの階差系列も分析し、どの系列同士の線形性が強いかどうかという観点から分析したうえでダミー変数を追加する [8]。この DAR はカオス時系列の予測において、場合によっては SETAR よりも有効に機能することが知られている [8]。

3.2 SETAR の残差を分析してダミー変数を設定する

そこで、本研究の「提案法②」として SETAR の残差の

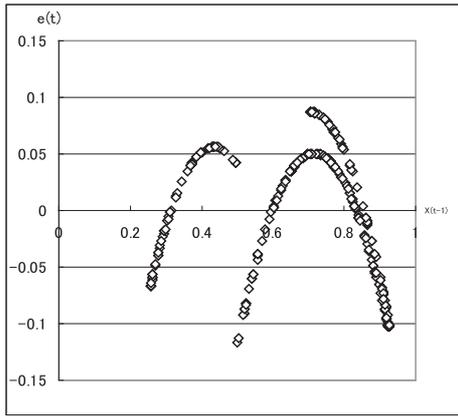


図 10 ロジスティック写像 1 の残差
Fig. 10 The error of the Logistic map 1.

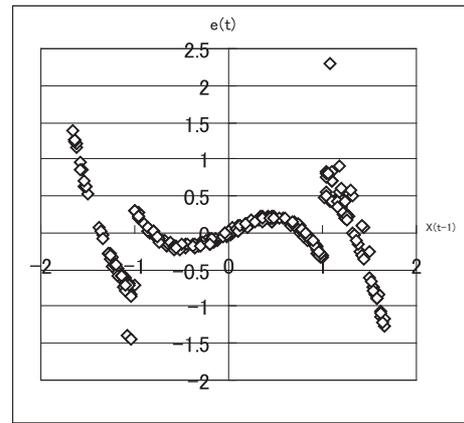


図 13 3 次の非線形写像の残差
Fig. 13 The error of the third-order nonlinear map.

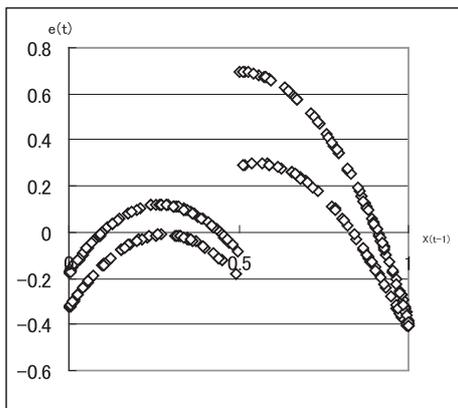


図 11 ロジスティック写像 2 の残差
Fig. 11 The error of the Logistic map 2.

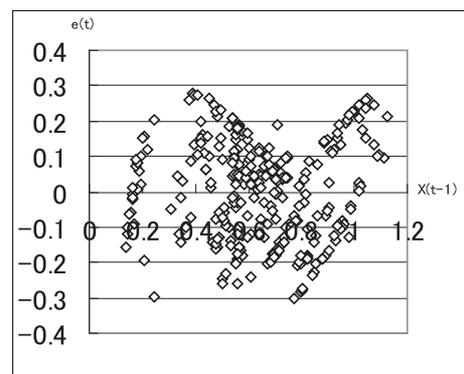


図 14 池田写像の残差
Fig. 14 The error of the Ikeda map.

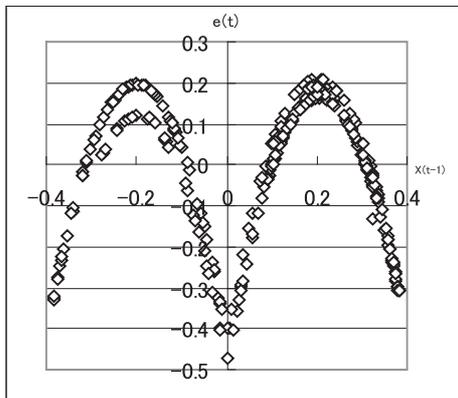


図 12 エノン写像の残差
Fig. 12 The error of the Henon map.

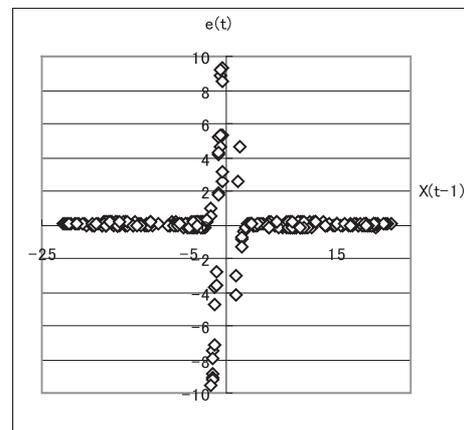


図 15 指数型 AR モデルの残差
Fig. 15 The error of the Ecp-AR model.

分析を行い、ダミー変数を設定することを考える。SETAR すでに大まかな非線形写像を近似した後の残差を分析しダミー変数を設定することにより、通常の DAR よりも精度の向上、そしてダミー変数の設定が容易になることが期待される。

ダミー変数を設定するには、やはり図を目視で判断せざるをえない。そこで①～⑥の各々のカオス時系列を従来法の SETAR で予測した後の残差を「縦軸：残差 $e(t)$ 、横軸：

$X(t-1)$ 」としてプロットしたものが図 9、図 10、図 11、図 12、図 13、図 14、図 15 である。⑦のテント写像に関しては、図は省略するが残差のオーダー（縦軸の最大値）が十分に小さく、式 (12) で付加している計算機イプシロンと同程度のレベルなので、これ以上は改善不要と判断した。

ダミー変数設定の基本的な方針は単純である。残差が平均的に 0 でない箇所、すなわち『残差の散らばりが ± に均等でない箇所』はまだ説明が見つことが期待されるのでそ

表 5 ダミー変数の設定

Table 5 The setting of the dummy variables.

ダミー変数の設定	ダミー変数①	ダミー変数②	ダミー変数③	ダミー変数④	ダミー変数⑤
ロジ3.7	$X(t-1) < 0.3$	$0.3 < X(t-1) < 0.5$	$0.5 < X(t-1) < 0.6$	$0.6 < X(t-1) < 0.85$	$0.85 < X(t-1)$
ロジ4.0	$X(t-1) < 0.15$	$0.15 < X(t-1) < 0.4$	$0.4 < X(t-1) < 0.5$	$0.5 < X(t-1) < 0.85$	$0.85 < X(t-1)$
エノン画像	$X(t-1) < -0.3$	$-0.3 < X(t-1) < -0.1$	$-0.1 < X(t-1) < 0.1$	$0.1 < X(t-1) < 0.3$	$0.3 < X(t-1)$
3次画像	$X(t-1) < -1.5$	$-1.5 < X(t-1) < -1$	$-1 < X(t-1) < 1.5$	$1.5 < X(t-1)$	
池田画像	$X(t-1) < 0.2$	$1 < X(t-1)$			
指数AR	$-2 < X(t-1) < -1.3$	$-1.3 < X(t-1) < 0$	$0 < X(t-1) < 1.3$	$1.3 < X(t-1) < 2$	
テント画像	-	-	-	-	-

のような箇所にダミー変数を設定する。ゆえに『残差 = 0』の横軸をデータの集合がクロスする箇所がダミー変数を設定する区間の「しきい値」の目安となる。加えて SETAR の際にもデータの性質がその地点で異なる「不連続点」を用いたが、ダミー変数設定のための「しきい値」としては、テント画像の中央ような互いに接する「不連続点」ではなく、図 6 の $X(t-1) = 0.5$ 付近の互いに接しない「不連続点」を目安とする。テント画像のように互いに接している「不連続点」周辺では、残差も同じ程度なので分割する必要がないからである。

『残差の散らばりが±に均等でない箇所』にダミー変数を割り当てるために、

- (a) 『残差 = 0』の横軸をデータの集合がクロスする箇所
 - (b) 互いに接しない「不連続点」
- という『2つの方針』に従い「しきい値」を選びダミー変数を以下のように設定した。

ロジスティック画像 1 については、図 10 により横軸とクロスする点は $X(t-1) = 0.3, 0.6, 0.85$ の周辺である。加えて $X(t-1) = 0.5$ で明らかに構造が異なる。これは不連続点である。以上の「しきい値」をもとに区間を定めたのが表 5 の 1 行目である。

ロジスティック画像 2 については図 11 により $X(t-1) = 0.15, 0.4, 0.9$ の周辺がクロスする点で、加えて同様に $X(t-1) = 0.5$ で明らかに構造が異なる。また、 $[0.6 \sim 0.9]$ の区間は変位が大きいので、さらに複数の区間に分割を行う方が精度を改善させる可能性があるのだが、あまり細かく分割を行うと恣意性が強くなるので、ここでは行っていない。同様にして各種のカオスを従来法の SETAR で予測した残差をもとにダミー変数を設定した一覧が表 5 である。

池田画像ではダミー変数を設定する区間が少なく、設定された区間も連続ではないが、これは図 14 において $[0.2 \leq X(t-1) < 1]$ の区間はおおむね『残差の散らばりが±に均等』であるのでダミー変数を設定しても効果がないからである。厳密に言えば、 $X(t-1) = 0.4$ 付近では残差が若干プラスに偏っているが、これについてもあまり細かい設定を行うと恣意性が強くなるのでここでは行っていない。

また注意すべきは、指数型 AR モデルに関して図 15 では $X(t-1) = 0$ 付近の構造がよく確認できないため、

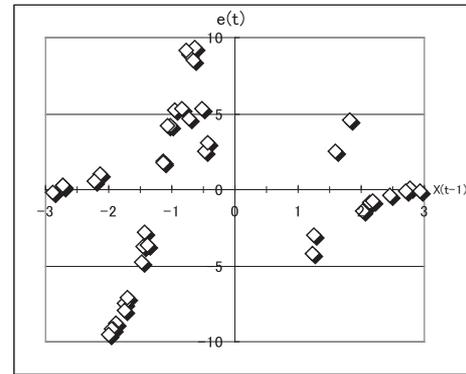


図 16 指数型 AR モデルの残差 (拡大)

Fig. 16 The error of the Exp-AR model (enlargement).

その周辺を拡大したのが図 16 である。これをもとに $X(t-1) = -2, -1.3, 0, 1.3, 2$ による区間にダミー変数を設定した。

この表 5 をもとにして文献 [8] と同様にたとえば $[X(t-1) < 0.3]$ で「ある : 1, 「ない : 0」とったダミー変数を元の SETAR に加えて再推定を行う。つまり式 (5) に $[a'_0 D + a''_0 D_2 + a'''_0 D_3 + \dots + \sum_{i=1}^K a''''_i D_3 X(t-i) + \dots]$ といった形でダミー変数を必要な数だけ追加して、再度「SETAR + ダミー変数」で予測を行うのが「提案法②」である。「提案法②」を用いた際の MSE が表 1 の一番右側であるが、通常の SETAR よりも精度が改善され、7 種類のカオス時系列において安定的な予測が行えることが分かる。加えて、必要以上に局所化した「提案法①」と違い、必要な箇所にだけ局所的なダミー変数の処理をした「提案法②」は、指数型 AR モデルにおいてもオーバフィッティングの兆候が見られないことが確認できる。また、ダミー変数の数が表 5 の程度では行列が特異になることはないので「提案法②」に関してはパラメータの推定に最小 2 乗法を用いている。

4. 「ファーマーとシドロウィッチの方法」と FAR によるベンチマーク

4.1 「ファーマーとシドロウィッチの方法」と FAR

本研究の時系列の予測においては、文献 [8] と同様にベンチマークとして「ファーマーとシドロウィッチの方法」[3] を簡略化した FAR を用いて精度の比較検証を行う。以下

にその「ファーマーとシドロウィッチの方法」と FAR の概要を述べる。

カオス時系列予測法の中でも「局所近似」の代表的な手法として「ファーマーとシドロウィッチの方法」が知られている。この方法では、状態空間内におけるカオスアトラクタの軌道を考え、この軌道上の 1 点を $x(T)$ とする。この $x(T)$ の近傍を状態空間内における過去の値のデータベース中より探索し近傍郡 $x(k_n)$ を作成する。そして近傍郡 $x(k_n)$ と $x(k_n + 1)$ の関係を

$$X(k_n + 1) = A \cdot X(k_n) + B \quad (16)$$

といった具合に線形近似により推定し、

$$X(T + 1) \cong A \cdot X(T) + B \quad (17)$$

のように予測する方法である。これはファーマーらの手法において 1 次近似と呼ばれるもので、他にも 0 次近似、2 次近似等も同様に考えられるが通常「ファーマーとシドロウィッチの方法」といった場合は上記の 1 次近似を指す場合が多い [9], [30]。また、この際に必要となるパラメータは、「状態空間の埋め込み次元： m 」と「遅れ時間： τ 」、近傍郡の「点の数： n 」かあるいは「半径： r 」が必要となり予測式の推定には最小 2 乗推定が行われる。これらのパラメータの設定のうち、「埋め込み次元」については、ターケンスの埋め込み定理によりカオス時系列のアトラクタ次元を d とすれば、 $m \geq 2d + 1$ であれば埋め込み可能であることが知られている。「遅れ時間： τ 」については多くの設定基準がある [31] が、特にアトラクタ次元 d の推定については、GP 法 (Grassberger-Procaccia algorithm) が最も一般的である。しかし、この GP 法は非常に膨大な演算量がかかり多くのデータ量が求められる。このように「ファーマーとシドロウィッチの方法」は予測の前処理での演算量が膨大であることに加え、「適切な」次元の推定にかなりのデータ数が要求されるなど運用の面では欠点がある。瞬時に大量の品目の需要予測を行うようなきわめて実務的な場面においては、往々にしてデータ数も少ない。また現実の経済時系列データなどでは、次元を正確に推定することも簡単ではない [32], [33]。そういった場面に適応できる「簡便法」としてはこの予測法を用いることは難しい。上記の観点から「ファーマーとシドロウィッチの方法」そのものは SETAR や DAR のなどの「簡便法」を評価する際の比較対象としては有用ではないが、文献 [5] のようにこの手法を簡略化した FAR を用いた評価は有用であると思われる。

「ファーマーとシドロウィッチの方法」において $\tau = 1$ と考えると、領域の数が非常に多い SETAR と等価であることが知られている (たとえば文献 [9] の p.159)。本研究では文献 [8] と同様に $m = 2$, $\tau = 1$ と設定した局所的な「AR モデル」を FAR とする。 m が小さいとターケンスの埋め込み定理からすれば埋め込める保証はないが、「局所

AR モデル」と考えれば「簡便法」としては十分に通用するはずである。また、本研究の目的からも $\tau = 1$ とする「AR」タイプのモデルとの比較が望ましい。

4.2 FAR をベンチマークとした比較検証

$m = 2$, $\tau = 1$ と設定した局所 AR モデルである FAR による予測結果の MSE が表 6 である。BasicDataSet : 300 をデータベースとして NewDataSet : 300 の予測を行った結果である。FAR [30] は、局所近傍の 30 点から局所 AR モデルを推定したことを意味する。ほとんどのカオス時系列に対し、非常に小さい MSE を得ることができたが、指数型 AR モデルにおける精度が思わしくない。特に表 1 の線形 AR モデルとほぼ同じ精度になっていることが問題である。その他のカオス写像に関しては表 1 における「提案法②」と比較すると厳密に言えば FAR の精度が勝るが、「提案法②」もそれほど遜色のない値を示しており、指数型 AR モデルの精度をふまえて総合的に判断すれば「提案法②」の方が安定した予測法であるといえる。

4.3 FAR の精度劣化の原因

FAR は直観的には局所近似を利用した有効な予測法と考えられるが、指数型 AR モデルの予測に関しては、線形 AR モデルと同程度という予想外の結果に終わっている。これは文献 [8] でも同様の傾向が確認できる。この原因は以下に述べるとおりである。

指数型 AR モデルを AR モデルで推定した後の残差が図 15 であるが $X(t - 1) = 0$ 付近の値が非常に特徴的である。これらを元のカオスアトラクタ上の図 5 で確認することは容易ではないが、データの値を調べると実は図 6 におけるノイズのような部分が主にそれに該当している。また、図 15 の縦軸の残差をそのままに横軸を $X(t - 2)$ に差し替えたものが図 17 であるが、当該データ部分が大きな残差となっていることが確認できる。つまり、この特徴的

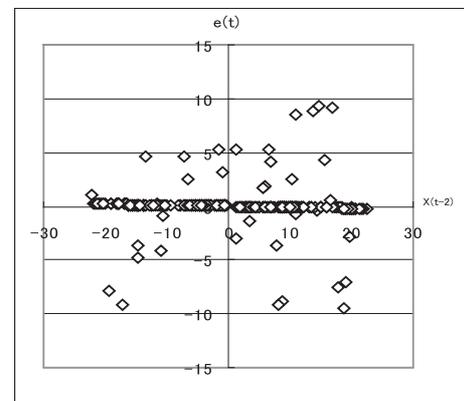


図 17 指数型 AR モデルの残差 (横軸 = $X(t - 2)$)

Fig. 17 The error of Exp-ARmodel (The horizontal axis = $X(t - 2)$).

なデータは $X(t-1)$ に関しては0付近であるが、 $X(t-2)$ に関しては非常に広範に広がっている。それゆえ、これらは $X(t-1)$ を基準にすればお互いが近傍と推計されるはずだが、 $X(t-2)$ を基準として考えると図 6, 図 17 で確認する限り近傍と推計される率が低いことが明白である。

指数型 AR モデルの場合、この特徴的な変動を首尾よく予測に反映させられるか否かが精度を左右する。ダミー変数を加える場合、残差分析によりダミー変数を設定することは表 5, 図 15 から分かる通り比較的容易な部類である。

ところが FAR の場合、 $m = 2$ とすると、ある点 $X(t)$ を予測しようとした場合に、 $\{X(t-1), X(t-2)\}$ のベクトルを基準として近傍群を作成し予測を行う。そのため、主に $X(t-1)$ の値と連動している図 15 のような特徴を処理することができない。

そこでこのような特徴をとらえるために、近傍の定義を $X(t-1)$ のみで行えるように FAR の次数を $m = 1$ に設定することを考える。この設定で各カオス時系列を予測した MSE が表 6 の左側であるが、今度は指数型 AR モデルに関して大幅に精度劣化が見られる。加えてその他のカオス時系列に対しても若干、精度が落ちている。この原因は $m = 1$ と設定すると近傍が $X(t-1)$ のみで定義されるが、AR の遅延次数も同時に 1 となるためである。「④ 3 次の非線形写像」や「⑥ 指数型 AR モデル」は式 (10) や式 (12) のモデルの中に $X(t-2)$ を含むため、AR の遅延次数が 1 では予測の精度が落ちるのは当然といえる。特に、指数型

AR モデルにおいては $X(t)$ と $X(t-1)$ の関係は図 5 のように線形の相関が見出せないが $X(t)$ と $X(t-2)$ の関係は図 6 のように非常にきれいな線形相関があり $X(t-2)$ の情報が落ちたことが大きく精度に影響を与えている。

結局、FAR で指数型 AR モデルを首尾よく予測するには、近傍の定義は前述の理由から $m = 1$ として $X(t-1)$ の近傍群を作成し、推定するモデルは $m = 2$ として局所的な AR(2) で予測する方法が良く、その結果が表 7 の右側である。この場合、指数型 AR モデルに関して FAR [20] は表 1 の「提案法②」に劣るが、FAR [10], FAR [30] は「提案法②」を上回っており、特に FAR [30] は本研究の中で最も良い精度を得ている。

しかし、その他のカオス時系列については表 5 に示した当初の FAR の方が精度が良い場合があるだけでなく、表 1 における「提案法②」の方が精度が勝るケースが「④ 3 次の非線形写像」に対する FAR [30], FAR [20] などで確認される。また、参考までに近傍を「データ数」ではなく「半径」で計算しても、表 6, 表 7 はほぼ同様の結果であった。

FAR は局所 AR モデルと考えればカオス時系列に対応しようと期待されたが、実際に運用においては指数型 AR モデルにおいて、近傍のノルム計算時の次数と予測の遅延次数を同一にすることが最適ではないケースが確認された。これはおそらく、ターケンスの埋め込み定理により保証されえないであろう $m = 2$ という小さい値を利用したこと起因すると思われるが、本研究ではカオスの相関次元解析等を行ったうえでの「ファーマーとシドロウィッチの方

表 6 FAR の精度

Table 6 The Accuracy of the FAR.

MSE	FAR[30]	FAR[20]	FAR[10]
ロジ3.7	0.000105589	2.3625E-05	2.01779E-06
ロジ4.0	0.000410079	9.87063E-05	8.65631E-06
エノン写像	0.015266586	0.008827956	0.002634593
3次写像	0.058883674	0.049280523	0.046668748
池田写像	0.008167139	0.006631923	0.005999718
指数AR	11.91235769	11.70349462	10.61295671
テント写像	0.001428646	0.000408135	0.000237926

表 7 パラメータ変更後の FAR の精度

Table 7 The Accuracy of the FAR after changing parameters.

MSE	AR 次数 = 1			AR 次数 = 2		
	近傍定義の次数 = 1			近傍定義の次数 = 1		
	FAR[30]	FAR[20]	FAR[10]	FAR[30]	FAR[20]	FAR[10]
ロジ3.7	0.000167035	6.33721E-05	3.00056E-05	0.000116498	5.44151E-05	2.66693E-05
ロジ4.0	0.000427423	0.000190951	6.01864E-05	0.0005918	0.000312971	0.000159551
エノン写像	0.095304358	0.078204777	0.061153873	0.044605967	0.033456216	0.025333454
3次写像	0.222598449	0.190549459	0.185754578	0.039838264	0.024002151	0.012982775
池田写像	0.023289675	0.021053254	0.021350745	0.010263423	0.008859037	0.010213445
指数AR	109.0537977	105.8829734	104.2554397	0.039838264	4.865379634	2.566418042
テント写像	0.00015304	0.000198154	0.000454263	3.58504E-05	0.000389904	0.000764564

法」は対象としていないため、この点の厳密な確認については別の機会に譲りたい。

5. 指数型 AR モデルと局所リアプノフ指数の分析

前章においては指数型 AR モデルに対する FAR の精度劣化に対し、アドホックな修正を加えることにより対応可能であることは示せた。しかし、本質的な理解や抜本的な対策のためには、もう少し議論が必要である。

本研究では7つのカオス時系列を生成するモデルを用いたが、指数型 AR モデルにおいてのみ、これほど予測に困難がともなうということは、やはり指数型 AR モデルが他に比べて異なる性質を持っていると考えるのが自然である。

指数型 AR モデルとその他の違いは、図 5 において $(x(t), x(t-1))$ 平面でのアトラクタがまったく1次元的不是なことである。つまり、 $x(t-1)$ の値によって、一意に $x(t)$ が決まらないということである。直観的にはこのことが、予測の難しさの原因ではないかと考えられる。しかしながら、図 6 においては $(x(t), x(t-2))$ 平面において、単純な線形回帰式があてはまる形状をしており、それらを総合的に考えれば $(x(t), x(t-1), x(t-2))$ の3次元上では、通常の重回帰分析で容易に推定される回帰平面で全体像をとらえることが可能であることが分かる。

また、指数型 AR モデルの誤差の状況 (図 15) からみても、多くの点においては誤差が0に近く、これは予測が成功しているということである。予測誤差を増大させているのは、 $x(t-1) = 0$ 付近の特定の点であることが明白である。

多くの点については回帰平面で説明ができ、特殊な異常値のようなものだけ性質が異なるということは、同じモデルの中でもこの特殊な点のみが非常にカオティックで残りの点はそれよりはスタティックであるということである。このようなカオス力学系の部分的な特徴を分析する際には「局所リアプノフ指数 (local lyapunov exponents)」が用いられる。

これまでの研究においてもカオス時系列の予測において、予測自体のほかに「局所リアプノフ指数」を用いた「予測精度の予測」などの研究が行われている。「予測精度の予測」が可能ということは同じ力学系から生成されるカオスであっても、地点によって予測しやすい場所とにくい場所が存在するということである。また、ロジスティック写像やローレンツモデルから生成される時系列データの予測において、この「局所リアプノフ指数」の情報を用いることによって「局所近似手法」よりも精度が向上する可能性があることがすでに報告されている [34]。指数型 AR モデルにおいて FAR による局所線形近似手法が、特定の地点のデータのみにおいて著しい誤差を生じているということ

は、この「局所リアプノフ指数」と関係があるのではないかと考えられる。

局所リアプノフ指数は、通常のリアプノフ指数から容易に導くことができる。通常のリアプノフ指数はカオスの特徴づける代表的な指標として知られているが、1次元力学系 $f(x)$ のリアプノフ指数の定義は、以下の式 (18) である。

$$\lambda = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{t=1}^N \log |f'(x)| \quad (18)$$

これは力学系が伸びるか縮むかの「データ数: N 」における平均的な傾向であり、この λ が正だと伸びることになり、つまりはカオスの可能性が高いといわれる。上の式からも分かるとおり、微少な差異が時間の経過にともないどの程度拡大するかという量を『長期間の時間スパンで平均化』した指標である。

また、多次元力学系の場合は以下のようにヤコビ行列を

$$J_t = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_k} \\ \frac{\partial f_2}{\partial x_1} & \vdots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_k}{\partial x_1} & \frac{\partial f_k}{\partial x_2} & \cdots & \frac{\partial f_k}{\partial x_k} \end{bmatrix} \quad (19)$$

J_t として、初期値 $x(0)$ からスタートしたヤコビ行列の N 回積行列を $M(x(0), N)$ としその固有値を $\sigma_i(N)$ とすると

$$\lambda_i = \lim_{N \rightarrow \infty} \frac{1}{N} \log \sigma_i(N) \quad (20)$$

がリアプノフ指数となる [35]。そして、1次元力学系の場合と同様に、このリアプノフ指数 λ_i の最大のものが正のとき、その力学系はカオスである可能性が高いと判定される。しかし、これは「ヤコビ行列の乗算回数: N 」に着目してみれば、あくまでも力学系全体の「平均的な」ダイナミクスを表す指標であり、このリアプノフ指数だけでは『各個別の点の周辺』に関して、非常にドラスティックな変化をするのか、カオスとはいえども比較的安定的な変動をするのかということは判断できない。そこで、議論されているのが局所リアプノフ指数である。局所リアプノフ指数は上記の式 (20) において、 $N \rightarrow \infty$ (実際に計算する場合は N は有限な大きな値で打ち切る) という力学系全体の性質を考えるのではなく、任意の有限ステップの L 期先で計算をとめてしまう。ある初期値 x から有限回 L の時点でのリアプノフ指数が「局所リアプノフ指数」といわれ $\lambda_i(x, L)$ などと表す。

文献 [36] では、エノン写像、池田写像、ローレンツモデルなどを分析対象として、① x に関しては平均化を行い時間的に $L = 2, 5, 10, 100, 1,000$ などと変化させた場合のリアプノフ指数の分布の推移、② 各 x における $L = 5, 1,000$ などとしたリアプノフ指数の分布を考察している。

本研究では、1期先の予測のみを対象としているので、 $L = 1$ の際の各点 x における状況を考える。 $L = 1$ と設定

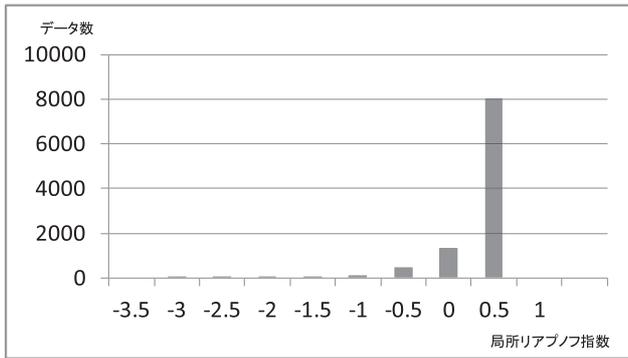


図 18 ロジスティック写像 1 の局所リアプノフ指数

Fig. 18 The local lyapunov exponents of the Logistic map 1.

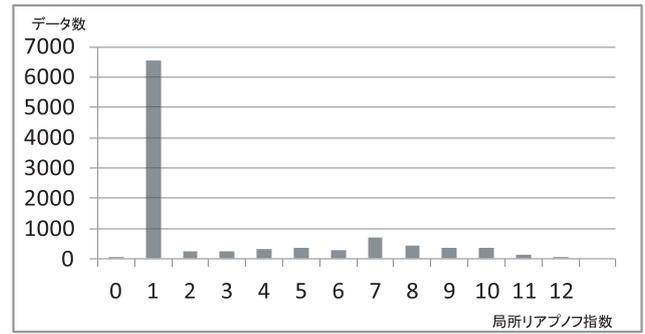


図 21 指数型 AR モデルの局所リアプノフ指数

Fig. 21 The local lyapunov exponents of the expAR model.

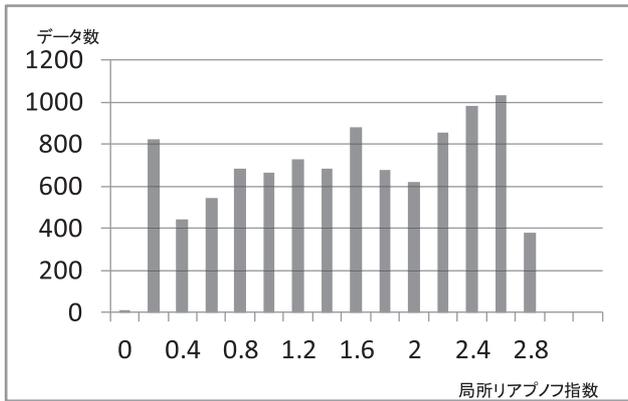


図 19 エノン写像の局所リアプノフ指数

Fig. 19 The local lyapunov exponents of the Henon map.

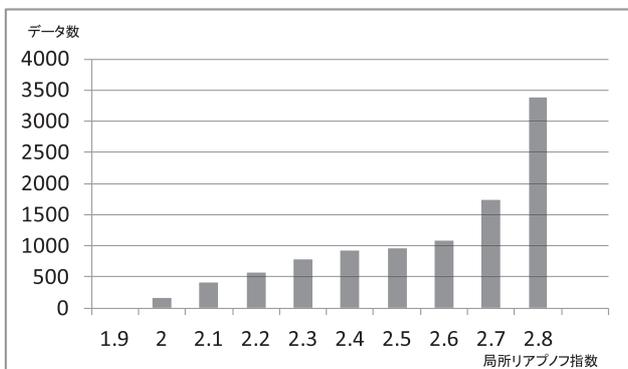


図 20 池田写像の局所リアプノフ指数

Fig. 20 The local lyapunov exponents of the Ikeda map.

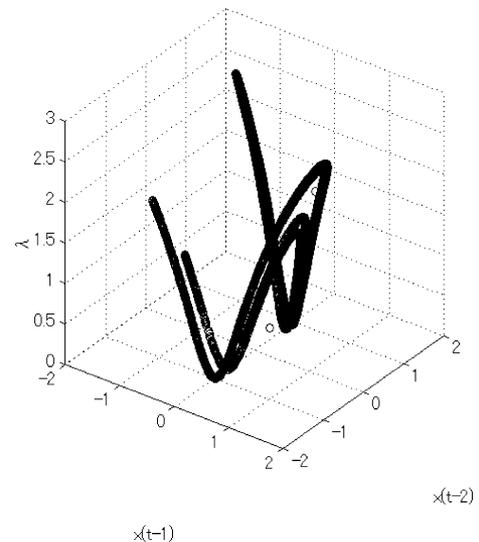


図 22 エノン写像の局所リアプノフ指数 (3次元散布図)

Fig. 22 The local lyapunov exponents of the Henon map (3D scatter diagram).

の低次に集中しており、力学系全体の平均的な傾向はカオスであっても、その中では比較的安定的な点が大多数を占めていることが分かる。しかしながら、非常に局所リアプノフ指数が高い地点が点在することが、グラフから読み取れる。

図 18 のロジスティック写像では、局所リアプノフ指数が「0~0.5」に集中しておりカオスの中でも比較的安定的な点が多数だが、局所的には、さらに局所リアプノフ指数が負である安定的な地点が点在することが分かる。それに対し、図 19、図 20 のエノン写像、池田写像においては、局所リアプノフ指数が高い箇所と低い箇所が比較的均一に分布していることが分かる。

また、局所リアプノフ指数 λ_i を縦軸として $x(t-1)$ 、 $x(t-2)$ とともに 3次元上に描いてみると図 22、図 23、図 24 のようにさらに異なった情報を得ることができる。エノン写像、池田写像においては、先の 1次元のヒストグラムにおいて局所リアプノフ指数の高低が各種混在していることが確認されたが、3次元上の図 21、図 22 で確認した場合、各点 x に関してその値は滑らか連続的に変化する

した場合は結局、ヤコビ行列自身の最大固有値で力学系の挙動を分析することになる。 L を大きくするとヤコビ行列のイタレーションの際に QR 分解などを用いなければ数値的に不安定になるが $L = 1$ ではそれも不要となる。

$L = 1$ のときの局所リアプノフ指数 $\lambda_i(x, 1)$ 、すなわちヤコビ行列の最大固有値のヒストグラムをいくつかの写像のデータにて描いたのが、図 18、図 19、図 20、図 21 である。縦軸がデータ数、横軸が局所リアプノフ指数であり、 x は 10,000 点を用いた。図 21 の指数型 AR モデルは、予想どおり局所リアプノフ指数が正でありながらも「0~1」

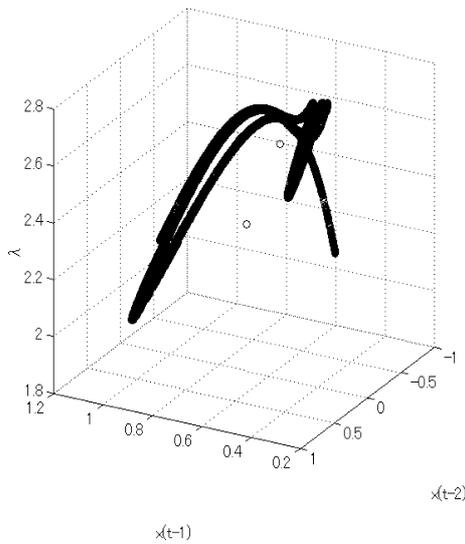


図 23 池田写像の局所リアプノフ指数 (3次元散布図)

Fig. 23 The local lyapunov exponents of the Ikeda map (3D scatter diagram).

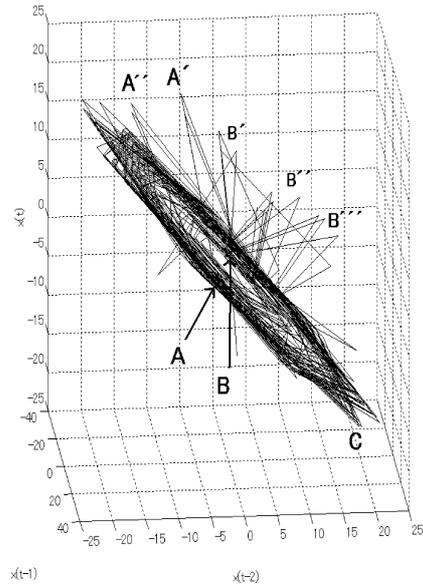


図 25 指数型 AR モデルの 3次元上でのアトラクタ

Fig. 25 The Attractor of the expAR model (3D scatter diagram).

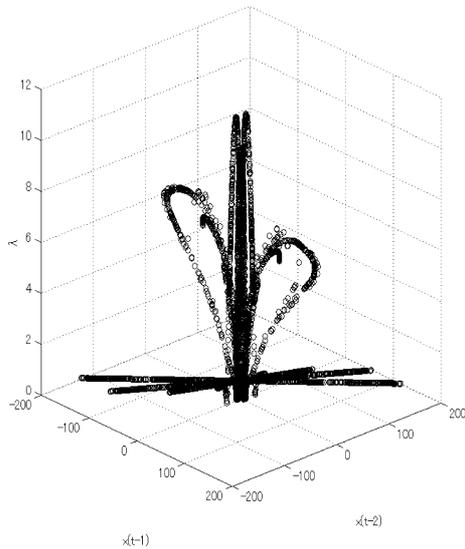


図 24 指数型 AR モデルの局所リアプノフ指数 (3次元散布図)

Fig. 24 The local lyapunov exponents of the expAR model (3D scatter diagram).

ことが分かる。ところが、図 23 の指数型 AR モデルは、これらとは明らかに性質が異なり、 $x(t-1) = 0$ 付近で突出して局所リアプノフ指数の値がはね上がっていることが分かる。連続的ではあるが変化が急峻である。このことが、FAR での予測が指数型 AR モデルにおいてのみ誤差を増大させる原因となっていることが分かる。

ここで注意しなければならないのは、局所リアプノフ指数の値が大きいから誤差が増大しているわけではないことである。確かに力学系の平均値としてのリアプノフ指数は大きい方が先々の挙動が乖離するスピードが速いゆえに予測が難しくなる。しかしながら、本研究で扱っているような局所化した、あるいは細分化したモデルによる 1 期先予測の場合、局所リアプノフ指数が大きくても、その点 x の

近傍が一様に同程度の局所リアプノフ指数を持っていれば、局所近似は可能だと考えられる。ここでは図は割愛するが、局所リアプノフ指数と FAR の予測誤差を同様に 3次元にプロットしても相関関係は見出せない。しかしながら、点 x の近傍の局所リアプノフ指数の変化が急峻である場合、その近傍は個別ごとに値が大きく異なる局所リアプノフ指数を有する集合になる可能性が否めない。

このことが指数型 AR モデルの予測を難しくしている。指数型 AR モデルは、ヤコビ行列を求めることができる連続関数であるが、『 $40 \exp(-X^2(t-1))$ 』の部分において $X(t-1)$ に非常に小さな値が入力された際にパラメータの大きさも相まって、連続関数から急激に突出した変化が出現するということが予測の難しさになっているということになる。このような性質を持つ関数から、離散型の時系列データが生成された場合、不連続点を持つ関数が当該不連続的近傍において、他の滑らかな箇所に比べて、値域に大きな変動を生じうることと類似の変動が生じている状況になる。

結局、指数型 AR モデルのアトラクタを 3次元にプロットしてみると図 25 のようになる。この図を説明すると、斜めの円軌道をぐるぐる回っていた点が、何かのタイミングで点 A のような $X(t-2) = 0$ 付近の軌道に入った場合、1 時刻進むと上記の A' や A'' へと大きく円軌道から逸脱する。あるいは B から B' , B'' へと逸脱する。そして、これらの A' , A'' , B' , B'' がすべて $X(t-1) = 0$ 付近の軌道上にのっているわけである。また、例外的に C などの点も推定される回帰平面より下へ軌道を逸脱しているが、これも $X(t-1) = 0$ 付近となっている。

指数型 AR モデルの特性を理解するために「局所リアプ

ノフ指数」の分析を行ってきたが、ここでははっきりしてきたことは、局所近似手法において、「不連続点」のような値域の変動が急峻になりうる点の周辺では、近傍のいくつかの点の挙動を鵜呑みにして推定を行うと大きな誤差を生む可能性があるということである。それゆえ、「局所近似手法」はこれらを想定した改良の余地が残されている。

そして、ダミー変数を用いた「残差分析」は、そういった特定のポイントだけを、いわば応急処置するのに最適な方法であったといえる。

6. まとめ

本研究ではカオス時系列予測問題における SETAR の性能評価とその改善を行った。従来の SETAR では、「ラグ」や「しきい値」の設定に分析者の主観的な判断が介在していた。またさらに精度としても改善の余地があることが本研究において確認された。そこで、区分数を増加させた細かい局所 AR モデル (提案法①) としての SETAR を用いれば、区分化する際の恣意性が排除され精度も向上することが期待された。しかしながら、実際には精度が劣化するケース、あるいはオーバーフィッティングの兆候を示すケースなどが存在した。それゆえ、SETAR にダミー変数を加えるという DAR の手法を応用したモデル (提案法②) により精度の改善を図り、こちらは首尾よく精度を改善することができ、本研究において扱ったカオス時系列においてはオーバーフィッティングの兆候を示す例も確認されなかった。

また、SETAR の性能評価以外にもいくつかの知見が得られた。SETAR の区分数が増加しても「係数ダミー」を用いて式を書き直すことにより必要な最小 2 乗法は 1 回となり区分数が増加してもムーアペンローズ型一般逆行列を用いることにより安定的な処理をすることが可能であった。

次に、カオスの構造の分析においてカオス写像を局所的に分割すると部分的に存在した線形の相関構造がかえって崩れるケースが存在することが判明した。

そして、カオスの予測法として「ファーナーとシドロウィッチの方法」を簡略化した FAR をベンチマークに用いたが、単に簡略化された FAR では指数型 AR モデルにおいて線形 AR モデルと同程度の精度しか望めず、FAR の運用においては近傍の定義と局所 AR モデルの次数は必ずしも同一が最適であるとは限らないことが判明した。

さらに付け加えれば、AR モデルにダミー変数を加えた DAR は文献 [8] によれば通常の SETAR と比べ精度としてほぼ同程度であることから、DAR の運用としても通常の AR モデルよりも SETAR にダミー変数を加えた方が精度が向上し、かつ階差などをとる必要も生じないことが分かる。

本研究では、ベンチマークに AR モデル、FAR を用いた。AR モデルは最小 2 乗法が 1 回で済むが、FAR は予測するデータセットの数が n 個なら最小 2 乗法も n 回必要となっ

てくる。それに対し、SETAR、提案法① (SETAR(T)) は精度が向上するにもかかわらず最小 2 乗法は 1 回で済む。提案法② (SETAR+ダミー) は、1 度 SETAR を推定した後、残差分析をして、さらにもう 1 度最小 2 乗法が必要なので計 2 回である。計算量の面においても FAR よりは提案法①、②の SETAR 系モデルの方が優れていることが分かる。

最後に、予測が難しい指数型 AR モデルの理解のため、「局所リアプノフ指数」による分析を行った。局所リアプノフ指数に基づく力学系の動的特性のとりえ方・考え方について、日本語では文献 [37] の第 4 章などが比較的平易でインターネット上で誰もが閲覧可能である。

この分析の結果、局所線形近似手法は「不連続点」のような値域の変動が急峻な点の周辺における運用に十分注意を払うべきだということが分かった。この効果的な運用が今後の局所線形近似手法の課題である。また、提案法①のような機械的処理に適応できる可能性を持った予測法は産業上の利用可能性が高い。細分化することにより、相関構造が崩れてしまうことにも対応可能な予測法を検討したい。加えて、提案法②に関しても自動化可能な設定方法が開発されれば、FAR などよりよほど効率的予測が行える手法になると考えられる。

参考文献

- [1] 合原一幸：カオス学入門，放送大学教育振興会 (2001).
- [2] 合原一幸 (編著)：ニューロ・ファジィ・カオス，オーム社 (1993).
- [3] Farmer, J.D. and Sidorowich, J.J.: Predicting Chaotic Time Series, *Physical Review Letters*, Vol.59, No.8, pp.845-848 (1987).
- [4] Tong, H.: Non-linear Time Series - A Dynamical System Approach, pp.377-447, Oxford University Press (1990).
- [5] Chen, R. and Tsay, R.S.: Functional-Coefficient Autoregressive Models, *Journal of the American Statistical Association*, Vol.88, No.421, pp.298-308 (1993).
- [6] Chen, R. and Tsay, R.S.: Nonlinear Additive ARX models, *Journal of the American Statistical Association*, Vol.88, No.423, pp.955-967 (1993).
- [7] 長瀬隆久：AR モデルにダミー変数を加えたカオス時系列予測法，情報処理学会論文誌，Vol.43, No.10, pp.3247-3250 (2002).
- [8] 長瀬隆久：AR モデルを用いたカオス時系列予測法の再検討，情報処理学会論文誌，Vol.45, No.3, pp.747-760 (2004).
- [9] 松葉育雄：非線形時系列解析，朝倉書店 (2000).
- [10] 松葉育雄：ラグ回帰，しきい値モデル，カオスの臨海特性，電子情報通信学会論文誌 (A)，Vol.J81-A, No.3, pp.389-396 (1998).
- [11] 白砂堤津耶：[例題で学ぶ] 初歩からの計量経済学，pp.156-158，日本評論社 (1998).
- [12] 松田安昌：非線形回帰モデルによるヘドニック・アプローチ，季刊住宅土地経済，No.52, pp.29-35 (2004).
- [13] 縄田和満：C による統計データ解析入門，p.193，東洋経済新報社 (2001).
- [14] 本多正久：経営のための需要の分析と予測，p.73，産能大学出版社 (2000).

- [15] 後藤昌司 (著), 浅野長一郎 (監修): 多変量データの解析法, pp.75–86, 科学情報社 (1973).
- [16] 早川 毅: 回帰分析の基礎, pp.94–103, 朝倉書店 (1986).
- [17] 佐和隆光: 回帰分析, pp.167–169, 朝倉書店 (1970).
- [18] D.A. ハーヴィル (著), 伊理正夫 (訳): 統計のための行列代数 (下), p.191, シュプリンガー・ジャパン (2007).
- [19] 田辺国土, 一般逆行列 (1), オペレーションズ・リサーチ, Vol.21, No.4, pp.213–215 (1976).
- [20] 新田康平: 独習応用線形代数—基礎から一般逆行列の理工学的応用まで, p.66, 近代科学社 (2006).
- [21] 伊理正夫: 一般線形代数, p.261, 岩波書店 (2003).
- [22] 金江春植 (著), 楊子江 (訳), 和田清 (監訳): 信号処理のための線形代数, p.21, 森北出版.
- [23] 上田太一郎: データマイニング実践集, pp.162–168, 共立出版 (1999).
- [24] 永田 悠, 立蔵洋介, 猿渡 洋, 鹿野清宏: 音場再現システムにおける環境変化に適応的な逆フィルタの逐次的緩和アルゴリズム, 電子情報通信学会論文 (A), Vol.J86-A, No.8, pp.824–834 (2003).
- [25] 潮 俊光: カオス制御, p.75, 朝倉書店 (1996).
- [26] Ikeda, K.: Multipule-Valued Stationary State and Its Instability of The Transmitted Light by a Ring Cavity System, *Optics Communications*, Vol.30, No.2, pp.257–261 (1979).
- [27] 合原一幸, 相澤洋二 (編著): カオス研究の最前線—非線形科学の世紀へ向けて, p.101, サイエンス社 (1999).
- [28] 久米 均, 飯塚悦功: 回帰分析, p.19, 岩波書店 (1987).
- [29] 山田剛史, 杉澤武俊, 村井潤一郎: R によるやさしい統計学, p.322, オーム社 (2008).
- [30] 田中 賢, 王 麟元, 松葉育雄: 経済時系列データの構造変化の抽出と予測, 電子情報通信学会論文誌 (A), Vol.J81-A, No.4, pp.658–663 (1998).
- [31] 合原一幸 (編): カオス時系列解析の基礎と応用, 産業図書 (2000).
- [32] 寺崎 健, 池口 徹, 合原一幸, 田中 智: 経済時系列データの決定論的非線形ダイナミカル特性に関する解析, 電子情報通信学会論文誌 (A), Vol.52, No.4, pp.1601–1617 (1995).
- [33] エドガー・ピーターズ (著), 新田 功 (訳): カオスと資本市場, 白桃書房 (1994).
- [34] Guegan, D. and Leroux, J.: Forecasting Chaotic systems: The role of local Lyapunov exponents, *Chaos, Solitons and Fractals*, pp.2401–2404 (2009).
- [35] 鈴木昱雄: カオス入門, p.125, コロナ社 (2000).
- [36] Abarbanel, H.D.I., Brown, R. and Kennel, M.B.: Variation of Lyapunov Exponents on a Strange Attractor, *J. Nonlinear Sci.*, Vol.1, pp.175–199 (1991).
- [37] 上島 豊: 荷電粒多体系におけるリアプノフ指数と巨視的現象に関する理論的研究, 大阪大学博士学位論文 (1997).



長瀬 隆久 (正会員)

2002年東京理科大学大学院経営学研究科修士課程修了。同年旧総務省郵政研究所, 2003年(財)土地総合研究所。2009年同研究所退職。2012年中央大学大学院理工学研究科博士後期課程満期退学(情報工学専攻)。同年東京テレコム基礎研究所研究員。研究分野は統計学, 時系列解析, 信号処理, ソフトコンピューティング等。