

社会システム分析プログラム College Analysis による 実データへの適用可能性の検証

～ 時系列分析モデル ～

石丸 敬二

概要

教育分野での利用を目的に社会システム分析に用いられる様々な手法を統合化したプログラムCollege Analysis Ver.3.5*を用いて、時系列データ分析の適用可能性を検証した。時系列分析は時間の経過と共に変化する変数の変動予測を目的とする分析である。ここでは、変動の分解モデルと呼ばれる手法を用いて、データを「傾向変動1(近似モデル) + 傾向変動2(予測モデル) + 季節変動(振幅変動×周期変動) + 残差」に分解するモデルを利用している。

キーワード

College Analysis, 社会システム分析, 統計, 時系列分析

1. はじめに

時系列分析は時間的に変化するデータの過去の特徴を調べてモデルを作り、未来を予測する手法である。これまで様々なモデルが提案されてきたが、歴史的に知られた予測モデルとしては、過去の変動の平均で予測する差の平均法、過去のデータからの影響を指数的に捉える指数平滑法やブラウン法、過去の最も近い状態を参考にする最近隣法、過去のデータを説明変数とした重回帰分析を利用する ARIMA などが有名である。また特に周期性が認められ

* © 福山平成大学経営学部経営学科福井正康教授, <http://www.heisei-u.ac.jp/ba/fukui/analysis.html>

るデータに対しては、変動を傾向変動、季節変動、循環変動、残差に分解して予測する、変動の分解モデルが有名である。傾向変動は増加や減少などデータの大局的な振る舞いを表し、移動平均や線形近似などの近似モデルを用いることが多い。季節変動は周期が一定の変動であり、循環変動は一般に周期が変化する変動である。データの全変動からこれらを差し引いたものが残差である。

著者は今後の学生の教育のために、時系列分析についてのプログラム開発に参画した²⁾。その中で考えられたのが、上に述べた手法を組み合わせた以下のモデルである。

$$\text{データの全変動} = \text{傾向変動 1 (近似モデル)} + \text{傾向変動 2 (予測モデル)} \\ + \text{季節変動 (振幅変動} \times \text{周期変動)} + \text{残差}$$

ここで季節変動については振幅の変化を考えて、振幅変動と振幅が一定の周期変動に分解している。循環変動についてはまだ組み込まれていない。それぞれの変動の詳細については参考文献¹⁾に詳しいのでここでは省略する。

2. 時系列分析の手法

時系列分析のプログラムは、College Analysis と呼ばれる社会システム分析教育用システムの一環で、その実行画面を図 2.1 に示す。

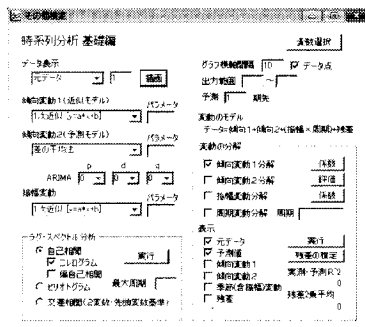


図 2.1 時系列分析実行画面

[変動の分解] グループボックス内の「傾向変動1」の分解、「傾向変動2」の分解、「振幅変動」の分解、「周期変動」の分解のチェックボックスをチェックするとそれぞれの分解が実行される。それぞれの分解の中で使う手法は左側のコンボボックスで指定する。

分解の実際を見てもらうために、図 2.2 のような時系列データのサンプルを考える。

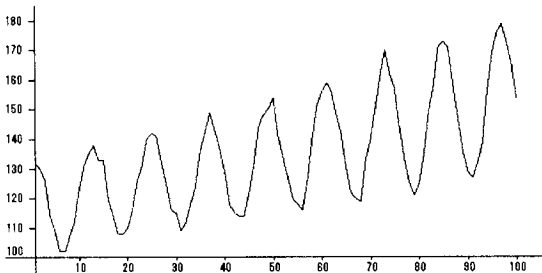


図 2.2 時系列データサンプル

傾向変動1 (近似モデル) には、移動平均法、1次近似、対数近似、べき乗近似、指数近似、多項式近似、一般の非線形近似などがある。例えば、2次式近似を用いると分解結果は図 2.3 のようになる。

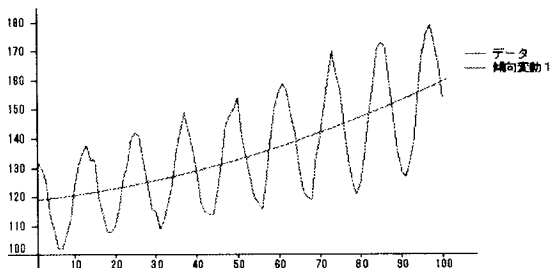


図 2.3 傾向変動の分解

傾向変動 2 (予測モデル) には、差の平均法、指数平滑法、ブラウン法、最近隣法、ARIMA などがあり、必要なパラメータは選択するコンボボックスの右隣や下にあるものを利用する。必要のない場合は入力ができなくなっている。このデータについて傾向変動 2 は考える必要がなさそうなのでここでは分解しないことにする。

傾向変動 1 を分解した残差に対して、季節変動の分解を行うが、まず振幅変動について考える。「振幅変動分解」チェックボックスの隣にある係数コマンドボタンをクリックすると、図 2.4 のように振幅変動の推定が行われる。1 次近似、対数近似、べき乗近似、指数近似、多項式近似がある中で、ここでは線形近似を用いている。

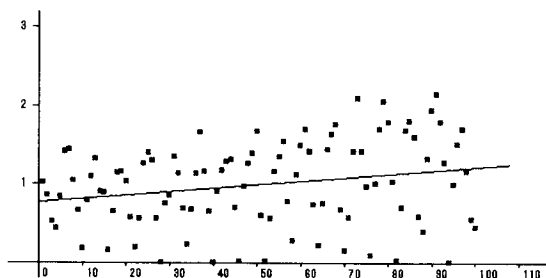


図 2.4 振幅変動の推定

振幅変動を除いたデータの振動数成分をピリオドグラムで見ると図 2.5 のようになる。

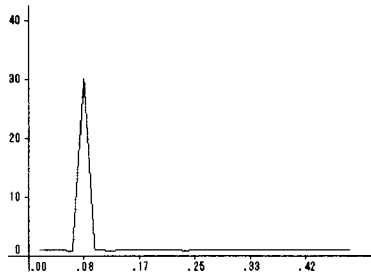


図 2.5 ピリオドグラム

これは周波数 0.083 (周期 12:これは図 2.5 のデータの数値表示から分かる) にピークを持っているので周期 12 の成分を周期変動として分解する。これらの分解を行うと最終的に図 2.6 のような結果となる。

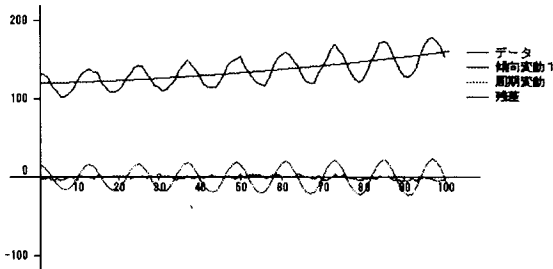


図 2.6 データの分解結果

残差はかなり小さくなっている。最後にデータとこのモデルでの予測値を図 2.7 に示す。

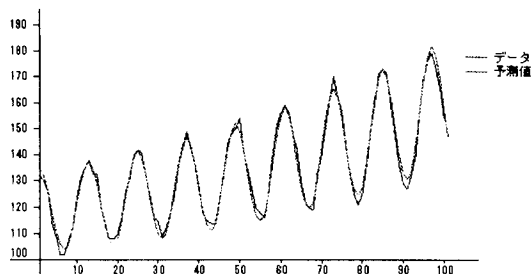


図 2.7 データと予測結果

これからもかなり精度よく予測が行われていることが分かる。予測結果の良し悪しを決める 1 つの指標である実測値と予測値の相関係数 R^2 値は 0.985 と高い値を示している。

実際のデータではなかなかこのようにうまくいかないが、分析の基本的な手順はここで述べた通りである。

3. 実データへの適用可能性

ここに実データによるいくつかのモデルを挙げ、分析の適用可能性を探ることとする。今回、社会科学系として日経平均株価のデータを、自然科学系として太陽黒点発生数のデータならびに広島県福山市における平均気温推移のデータをそれぞれ用いて検証してみた。

3.1 日経平均株価データ

データとして、1991 年 1 月から 2010 年 3 月まで、月単位の日経平均株価(終値) 231 期分を使用した。

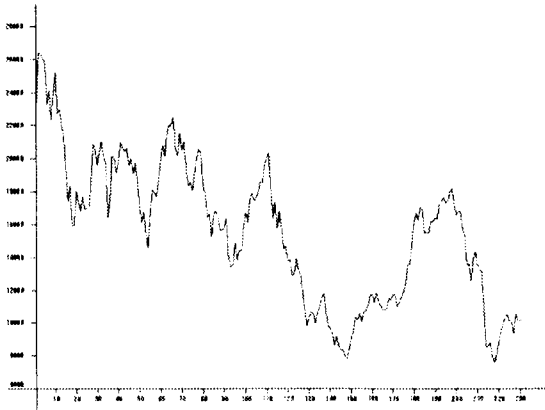


図 3.1 日経平均株価の時系列データ (1991.1 ~ 2010.3) ³⁾

3.1.1 予測モデルの比較

傾向変動 1 (近似モデル) の分解手法を「1 次近似」「対数近似」「べき乗近似」「指数近似」の 4 つ、傾向変動 2 (予測モデル) の分解手法を「差の平均法」「指数平滑法」「ブラウン法」「最近隣法」の 4 つにそれぞれ絞り込み、各組み合わせによる検定を行ってみた。その検定結果が次表である。

表 3.1.1 日経平均株価の予測モデルの比較

| No. | 傾向変動 1 | 傾向変動 2 | 傾向変動 2 パラメータ | R ² 値 | 残差 2 乗平均 (分散) | 標準偏差 |
|-----|--------|--------|-----------------|------------------|------------------|---------|
| 1 | 1 次近似 | 差の平均法 | — | 0.9474 | 1022060.16 | 1010.97 |
| 2 | 1 次近似 | 指数平滑法 | 0.99 | 0.9498 | 955564.37 | 977.53 |
| 3 | 1 次近似 | ブラウン法 | 0.71 | 0.9435 | 1209533.90 | 1099.79 |
| 4 | 1 次近似 | 最近隣法 | 1.62 | 0.8988 | 1744058.93 | 1320.63 |
| 5 | 対数近似 | 差の平均法 | — | 0.9463 | 1135985.14 | 1065.83 |
| 6 | 対数近似 | 指数平滑法 | 0.99 | 0.9447 | 1039074.20 | 1019.35 |
| 7 | 対数近似 | ブラウン法 | 0.51 | 0.9420 | 1107003.09 | 1052.14 |
| 8 | 対数近似 | 最近隣法 | 1.62 | 0.9078 | 1576402.59 | 1255.55 |
| 9 | べき乗近似 | 差の平均法 | — | 0.9423 | 1397855.96 | 1182.31 |
| 10 | べき乗近似 | 指数平滑法 | 0.99 | 0.9341 | 1239087.39 | 1113.14 |
| 11 | べき乗近似 | ブラウン法 | 0.52 | 0.9413 | 1142805.18 | 1069.02 |
| 12 | べき乗近似 | 最近隣法 | 1.62 | 0.9108 | 1524796.93 | 1234.83 |
| 13 | 指数近似 | 差の平均法 | — | 0.9474 | 1021778.88 | 1010.83 |
| 14 | 指数近似 | 指数平滑法 | 0.99 | 0.9498 | 955264.85 | 977.38 |
| 15 | 指数近似 | ブラウン法 | 0.51 | 0.9419 | 1092405.95 | 1045.18 |
| 16 | 指数近似 | 最近隣法 | 1.62 | 0.8915 | 1855510.95 | 1362.17 |

上表から、最も標準偏差の値が小さい（予測の誤差が少ない）のは、No.14の指数近似と指数平滑法（パラメータ 0.99）の組み合わせであることがわかる。また、予測の確からしさを表す R² 値を見ても、No.14 の値(0.9498) が最も高く、これらの予測結果の中では信用度が高いと言えるであろう。

3.1.2 振幅変動分解と周期変動分解

ここで、振幅変動分解を行ってみた。手法としては 1 次近似を用いている。

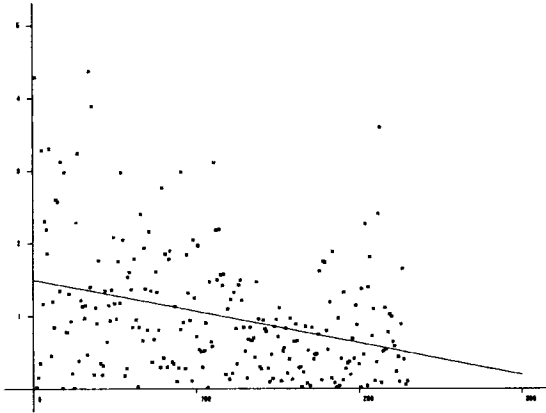


図 3.1.2a 日経平均株価の振幅変動の推定

振幅変動を除いたデータの振動数成分をピリオドグラムで見ると図 3.1.2b のようになる。

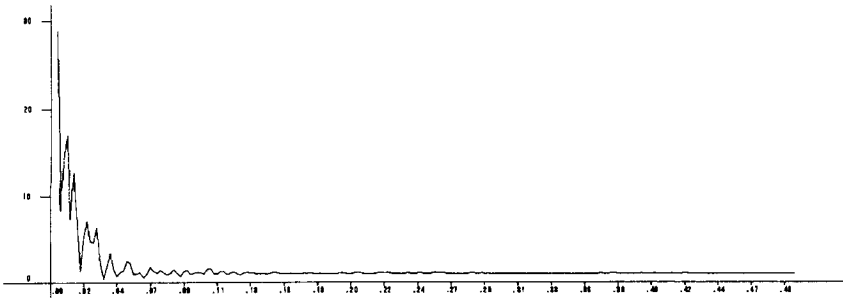


図 3.1.2b 日経平均株価のピリオドグラム

これは周波数 0.004 (周期 225.000) にピークを持っているが、データ期間 231 に対して有効な周期成分としては成り立たないため、周期変動分解は不可であった。

よって、このモデルに対しては周期変動分解を用いることはできず、傾向変動 1 および傾向変動 2 の分解により予測した結果、最適な予測モデルは図 3.1.2c となった。

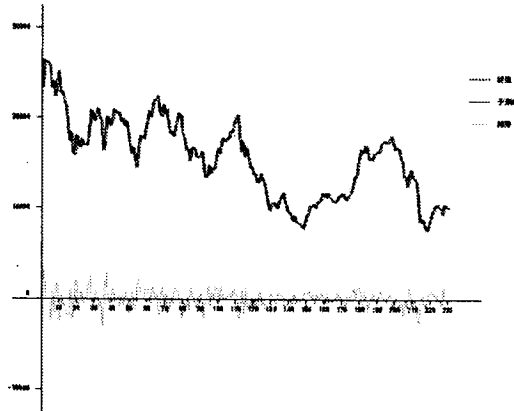


図 3.1.2c 日経平均株価の最適予測モデル

図 3.1.2c に示す予測値は、指数近似による傾向変動 1 の分解、および指数平滑法（パラメータ 0.99）による傾向変動 2 の分解によるものである。 R^2 値は 0.9498 を示し、残差の標準偏差は 977.38 であった。

ただし、実際の株価で考えた場合、1,000 円近くの予測誤差が出ているわけであり、現実に適用可能かと言えれば否であると言わざるをえないであろう。

3.2 太陽黒点数データ

データとして、1945 年 1 月から 1976 年 12 月まで、月単位の太陽黒点発生観測数 384 期分を使用した。

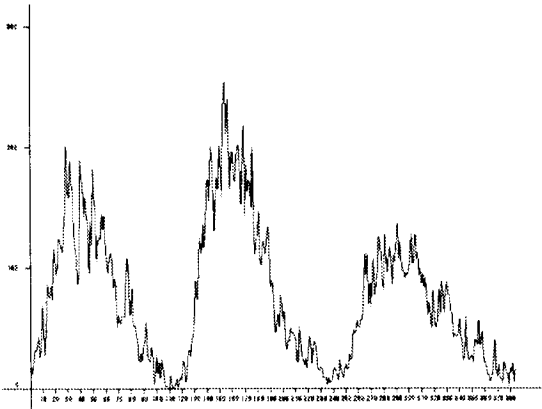


図 3.2 太陽黒点数の時系列データ (1945.1 ~ 1976.12) ⁴⁾

3.2.1 予測モデルの比較

前節同様、傾向変動 1 および傾向変動 2 の分解手法の各組み合わせによる検定を行った。次表が検定結果である。

表 3.2.1 太陽黒点数の予測モデルの比較

| No. | 傾向変動 1 | 傾向変動 2 | 傾向変動 2 パラメータ | R ² 値 | 残差 2 乗平均 (分散) | 標準偏差 |
|-----|--------|--------|-----------------|------------------|------------------|-------|
| 1 | 1 次近似 | 差の平均法 | — | 0.8896 | 368.71 | 19.2 |
| 2 | 1 次近似 | 指数平滑法 | 0.63 | 0.8987 | 327.73 | 18.1 |
| 3 | 1 次近似 | ブラウン法 | 0.24 | 0.8960 | 351.11 | 18.74 |
| 4 | 1 次近似 | 最近隣法 | 1.62 | 0.8096 | 619.98 | 24.9 |
| 5 | 対数近似 | 差の平均法 | — | 0.8897 | 369.23 | 19.22 |
| 6 | 対数近似 | 指数平滑法 | 0.63 | 0.8983 | 329.35 | 18.15 |
| 7 | 対数近似 | ブラウン法 | 0.24 | 0.8961 | 351.20 | 18.74 |
| 8 | 対数近似 | 最近隣法 | 1.62 | 0.8219 | 580.51 | 24.09 |
| 9 | べき乗近似 | 差の平均法 | — | 0.8897 | 369.48 | 19.22 |

| | | | | | | |
|----|-------|-------|------|--------|--------|-------|
| 10 | べき乗近似 | 指数平滑法 | 0.63 | 0.8981 | 330.06 | 18.17 |
| 11 | べき乗近似 | ブラウン法 | 0.24 | 0.8961 | 351.21 | 18.74 |
| 12 | べき乗近似 | 最近隣法 | 1.62 | 0.8205 | 585.47 | 24.2 |
| 13 | 指数近似 | 差の平均法 | — | 0.8896 | 368.73 | 19.2 |
| 14 | 指数近似 | 指数平滑法 | 0.63 | 0.8987 | 327.75 | 18.1 |
| 15 | 指数近似 | ブラウン法 | 0.24 | 0.8960 | 351.10 | 18.74 |
| 16 | 指数近似 | 最近隣法 | 1.62 | 0.8216 | 583.57 | 24.16 |

上表から、最も標準偏差の値が小さいのは、No.2 の 1 次近似と指数平滑法 (パラメータ 0.63)、および No.14 の指数近似と指数平滑法 (パラメータ 0.63) の組み合わせであることがわかる。また、 R^2 値を見ても、No.2 と No.14 が同値 (0.8987) で最も高く、これらの予測結果の中では信用度が高いと言えるそうである。

3.2.2 振幅変動分解と周期変動分解

ここで、振幅変動分解を行ってみた。手法としては 1 次近似を用いている。

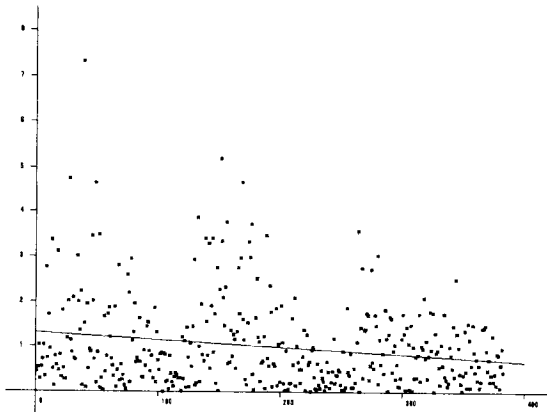


図 3.2. 2a 太陽黒点数の振幅変動の推定

振幅変動を除いたデータの振動数成分をピリオドグラムで見ると図 3.2.2b のようになる。

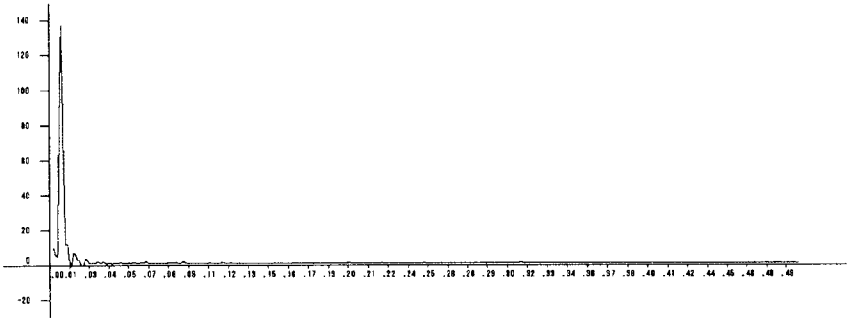


図 3. 2. 2b 太陽黒点数のピリオドグラム

これは周波数 0.008 (周期 126.000) にピークを持っており、データ期間 384 に対して、周期成分 126 による周期変動分解を行うこととする。

これらの分解、ならびに前項で導き出した傾向変動 1 および傾向変動 2 の組み合わせによる分解を行うと、最終的に図 3.2.2c のような予測モデルとなる。

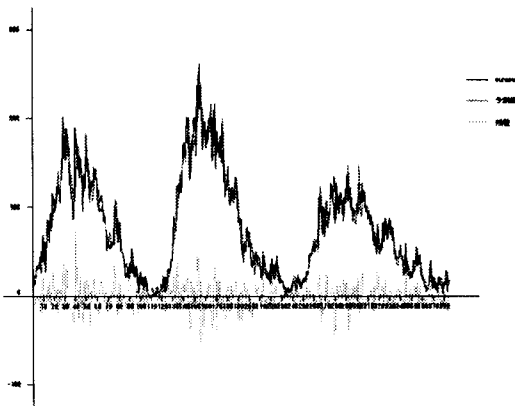


図 3. 2. 2c 太陽黒点数の最適予測モデル

図 3.2.2c に示す予測値は、1 次近似による傾向変動 1 の分解、および指数平滑法による傾向変動 2 の分解、さらに周期変動分解（周期成分 126）によるものである。R² 値は 0.9329 を示し、残差の標準偏差は 14.7 であった。傾向変動 1 の分解を指数近似に差し替えても結果は変わらなかった。

残差が小さくなったことから、このモデルに対しては周期変動分解を加えることは有効であったと言える。

3.3 平均気温データ

データとして、2005 年 1 月から 2009 年 12 月まで、広島県福山市における月単位の平均気温（摂氏）60 期分を使用した。

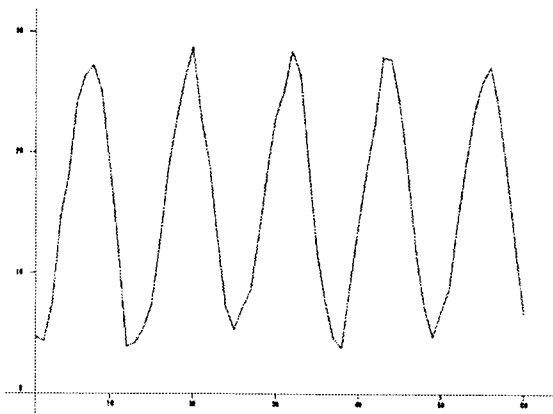


図 3.3 福山市の平均気温の時系列データ（2005.1～2009.12）⁵⁾

3.3.1 予測モデルの比較

前節同様、傾向変動 1 および傾向変動 2 の分解手法の各組み合わせによる検定を行った。次表が検定結果である。

表 3.3.1 平均気温の予測モデルの比較

| No. | 傾向変動 1 | 傾向変動 2 | 傾向変動 2 パラメータ | R ² 値 | 残差 2 乗平均 (分散) | 標準偏差 |
|-----|--------|--------|-----------------|------------------|------------------|------|
| 1 | 1 次近似 | 差の平均法 | — | 0.7235 | 21.27 | 4.61 |
| 2 | 1 次近似 | 指数平滑法 | 0.99 | 0.7198 | 19.92 | 4.46 |
| 3 | 1 次近似 | ブラウン法 | 0.99 | 0.9397 | 9.32 | 3.05 |
| 4 | 1 次近似 | 最近隣法 | 1.62 | 0.8465 | 12.11 | 3.48 |
| 5 | 対数近似 | 差の平均法 | — | 0.7230 | 21.15 | 4.6 |
| 6 | 対数近似 | 指数平滑法 | 0.99 | 0.7207 | 19.78 | 4.45 |
| 7 | 対数近似 | ブラウン法 | 0.99 | 0.9397 | 9.30 | 3.05 |
| 8 | 対数近似 | 最近隣法 | 1.62 | 0.8344 | 12.59 | 3.55 |
| 9 | べき乗近似 | 差の平均法 | — | 0.7232 | 21.15 | 4.6 |
| 10 | べき乗近似 | 指数平滑法 | 0.99 | 0.7208 | 19.78 | 4.45 |
| 11 | べき乗近似 | ブラウン法 | 0.99 | 0.9397 | 9.30 | 3.05 |
| 12 | べき乗近似 | 最近隣法 | 1.62 | 0.8235 | 13.96 | 3.74 |
| 13 | 指数近似 | 差の平均法 | — | 0.7235 | 21.28 | 4.61 |
| 14 | 指数近似 | 指数平滑法 | 0.99 | 0.7197 | 19.93 | 4.46 |
| 15 | 指数近似 | ブラウン法 | 0.99 | 0.9397 | 9.32 | 3.05 |
| 16 | 指数近似 | 最近隣法 | 1.62 | 0.8419 | 12.75 | 3.57 |

上表から、最も標準偏差の値が小さいのは、No.7 の対数近似とブラウン法（パラメータ 0.99）、および No.11 のべき乗近似とブラウン法（パラメータ 0.99）の組み合わせであることがわかる。また、R² 値を見ても、No.7 と No.11 が同値（0.9397）で最も高く、これらの予測結果の中では信用度が高いと言えそうである。

3.3.2 振幅変動分解と周期変動分解

ここで、振幅変動分解を行ってみた。手法としては 1 次近似を用いている。

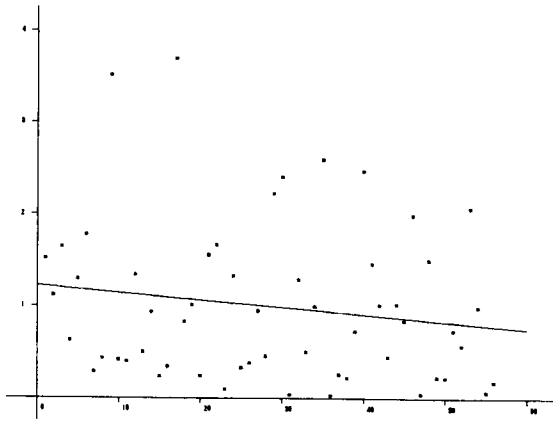


図 3. 3. 2a 平均気温の振幅変動の推定

振幅変動を除いたデータの振動数成分をピリオドグラムで見ると図 3.3.2b のようになる。

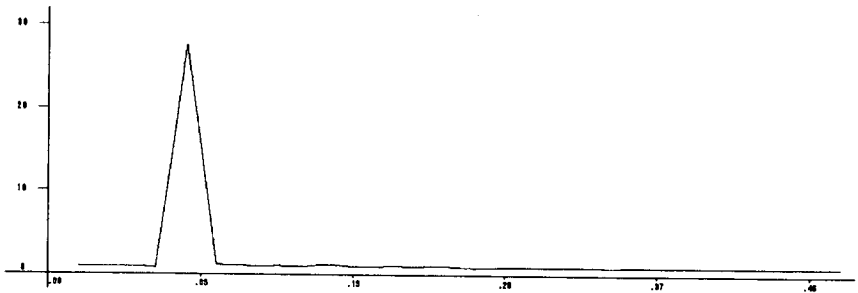


図 3. 3. 2b 平均気温のピリオドグラム

これは周波数 0.083 (周期 12.000) にピークを持っているので、周期成分 12 による周期変動分解を行うこととする。

これらの分解、ならびに前項で導き出した傾向変動 1 および傾向変動 2 の組み合わせによる分解を行うと、最終的に図 3.3.2c のような予測モデルとな

る。

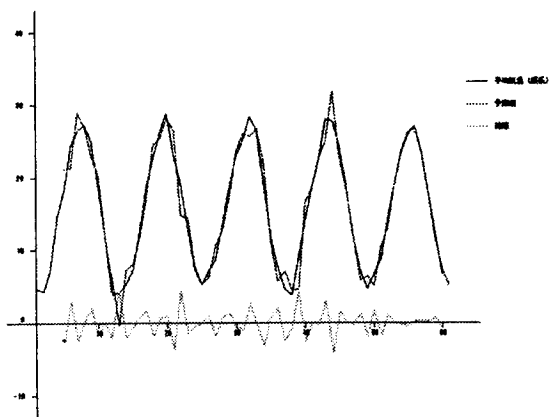


図 3. 3. 2c 平均気温の最適予測モデル 1

図 3.3.2c に示す予測値は、対数近似による傾向変動 1 の分解、およびブラウン法による傾向変動 2 の分解、さらに周期変動分解（周期成分 12）によるものである。R² 値は 0.9456 を示し、残差の標準偏差は 1.92 であった。傾向変動 1 の分解をべき乗近似に差し替えても結果は変わらなかった。

残差が小さくなったことから、このモデルに対しては周期変動分解を加えることは有効であったと言える。

3.3.3 周期変動分解のみによる検定

しかし、このモデルについては、前項よりも最適な予測モデルが存在する。傾向変動を排して、周期変動分解のみによる予測を行ってみた結果が図 3.3.3 である。

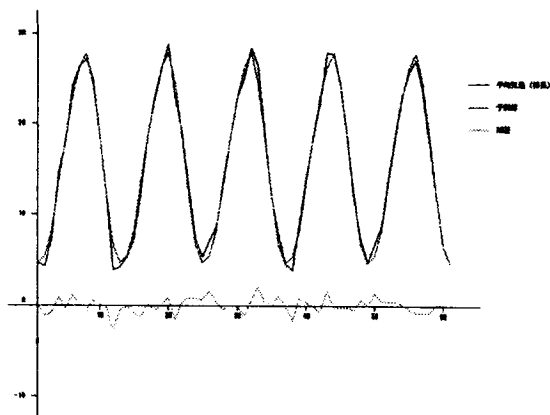


図 3.3.3 平均気温の最適予測モデル 2

図 3.3.3 に示す予測値は、周期変動分解（周期成分 12）のみによるものである。R² 値は 0.9893 を示し、残差の標準偏差は 0.71 であった。傾向変動分解を含めた場合よりも、予測誤差がかなり小さく抑えられた結果となった。四季の変化による影響を受けやすい気温データであることが原因と考えられる。

4. おわりに

今回、3つのデータモデルを用いて時系列分析を行ってみた。他にも試用したモデルは複数あったが、社会科学系と自然科学系とで結果が顕著なモデルに絞って取り上げたつもりである。見えてきたのは、周期変動の有無であった。社会情勢、政治、国際関係等、人心を含め種々の環境に左右されやすい株価については、周期変動分解による予測が極めて難しいことが実感された。それに対して、気温に代表される自然科学系データについては、突発的な環境変化がない限り、一定の周期によって変動している傾向が強い（湿度や降雨量等、気温ほど明確な周期変動が見受けられないデータも少なくはなかつ

たが)。

今後、検証データの種類を増やし、教材としても利用可能なモデルを少しでも多く用意していきたい。また、計測期間の長短で結果がどう変わるのか、データの種類によって最適な期間がどの程度なのかについても押さえておく必要がある。

最後に、今回の時系列分析をはじめ、社会システム分析プログラム“College Analysis”により各種統計処理の入口へと筆者を導いてくださった、福山平成大学経営学部経営学科福井正康教授に、深甚なる謝意を表します。

参考文献

- 1) 北川源四郎, 『時系列解析入門』, 岩波書店, 2005.
- 2) 福井正康, 王迎春, 王晶, 石丸敬二, 「社会システム分析のための統合化プログラム 11 - 時系列分析 -」, 『福山平成大学経営研究』, 6号, 2010.
- 3) 「Yahoo! ファイナンス」, Yahoo! JAPAN, 2010年3月26日,
<<http://finance.yahoo.co.jp/>>
- 4) 「Web Decomp」 統計数理研究所, 2010年1月6日,
<http://ssnt.ism.ac.jp/inets/inets_sample.html>
- 5) 「気象庁」, 国土交通省, 2010年1月7日,
<<http://www.jma.go.jp/jma/index.html>>

Verification of possible application to real data by social systems analysis program "College Analysis" - Time Series Analysis Model -

Keiji ISHIMARU

Department of Economics, Faculty of Economics,
Fukuyama University

Abstract

In this paper, for use in the field of education, by using integrated program for the analysis of social systems "College Analysis Ver.3.5", I verified the applicability of time series data analysis. The purpose of time series analysis is a forecast of the variables that change over time. I used a time series decomposition model as Trends1 (approximation model) + Trends2 (prediction model) + Seasonal variation (amplitude variation \times periodic variation) + Residual.

Keywords

College Analysis, social system analysis, statistics, time series analysis