

曜日効果の識別と確率モデル*

高 岡 慎†

On Statistical Modelling of Trading Day Effects in Economic Seasonality

Makoto Takaoka†

Many time series, such as economic statistics or government office statistics, are affected by “Trading-day effect” which is due to change of month length and composition of particular day in a month. In an analysis of a business condition trend, it is often desirable to remove trading-day effect from the original data in some way. In this paper, a method for detecting trading-day effect in monthly time series by spectral analysis and some stochastic models for trading-day effect in time domain are discussed. Then examples of the regression models with stochastically varying coefficients are given for real data, such as indices of industrial production of Japan.

経済統計や官庁統計などの月次または四半期の時系列データは、季節性の他に、1ヶ月に含まれる曜日の構成の変化や、月ごとの日数の違いによって生じる「曜日効果」(trading day effect,あるいはcalendar effect)を含む場合が多い。「曜日効果」は、データを用いた景気動向の分析などにおいては、季節性と同様、何らかの方法により元のデータから除去されることが望ましい場合が少なくない。本稿では月次データについて、スペクトル解析の手法を用い曜日効果の存在を識別する方法と、時間領域における確率モデルを用いた曜日効果のモデリングと調整について論じた。また従来方法を拡張した曜日効果の確率的係数モデルを提案し、鉱工業生産指数などのデータに対する適用を試みた。

1. はじめに

経済活動の多くは大きく1年というサイクルの中で運営されており、このことは経済時系列データの多くが季節性を帯びる一つの大きな要因となっている。1年のサイクルが生じる原因には、文字通り季節による気候の変化などの自然的要因や、慣習的要因などが考えられるが、いずれにせよ中長期的な経済動向の分析に際しては、季節性は適切な手法で除去されるのが望ましいとして、これまで数多くの季節調整の方法が研究されてきた。

一方季節性に類似しているものの、比較的注意されることの少ない問題として、曜日効果というものがある。これはカレンダーにおける日付と曜日の配置や休日の数などから生じる見かけ上の変動である。経済活動は1週間を1サイクルとする小さな周期があることが普通であり、これが曜日効果の主な原因となる。7日の周期とは具体的には毎週土曜と日曜が休日であったり特定の曜日が定休日であるといったことである。

曜日効果を含む時系列は四半期データや月次データなどがありうるが、本論では月次の時系

* 本論をまとめるにあたって、東京大学経済学部の国友直人先生、矢島美寛先生、竹村彰通先生、久保川達也先生、ならびにレフェリーの方より貴重なコメントを頂きました。ここに感謝の意を表します。

† 東京大学大学院経済学研究科, 〒113-0033 東京都文京区本郷 3-7-1

列のみが入手可能な場合に、そこから曜日効果の有無を判断し必要に応じて調整を行うための枠組みについて述べる。

集計データへの曜日効果の現れ方はデータがフローであるかストックであるかによって多少異なるが、主に以下の様な原因で現れる。まず(1)一ヶ月の長さが月によって異なるということ、(2)毎年曜日が一日ずつずれていくということ、(3)祝祭日の影響などである。(1)を原因とする変動はデータを月次集計すれば常に生じるもので、集計前のデータが1週間の周期を持っているかどうかとは直接関係が無い。また閏年を無視すればその周期はほぼ12ヶ月であり、季節性と混同されることが多い。日次の系列がもつ周期性は(2)との関連で集計データに周期変動をもたらす。カレンダー上の日付と曜日の年ごとのずれはある月に関して前年同月と比較すると営業日が1日多いまたは少ないという事態を引き起こす。こうした変動は経済動向を反映するものではないので(3)によって生じる効果とともに、必要に応じて何らかの調整が為されることが望ましい。

季節調整においては調整の適切さを判断する際に、データのスペクトルを推定し、季節周波数($1/12 \div 0.082 \text{ cycle/month}$)とその整数倍の周波数におけるピークもしくはディップの有無を確認するという方法がある。曜日効果に関してもその周期性に注目することにより、季節性と同様に周波数領域においてその存在を識別する方法がCleveland-Devlin [1] などにより提案されている。月次データから見かけ上12ヶ月の周期的変動が除去されればそれでよい季節調整とは異なり、その取り扱いはやや複雑となるが、データから推定されたスペクトルが扱う時系列が曜日効果を含むかどうかのひとつの判断基準を与える。

一方、単にデータに調整を施して全体的なトレンドを見るだけでなく、データを季節性などのいくつかの要素に分解して分析を行ったり、モデルによる予測を試みる場合は、曜日効果も一要素として明示的なモデルによって取り扱った方が、より精度の高い結果が得られることが期待できる。

以下次節ではフローデータに関する曜日効果の周期的特性についてまとめ、第3節では実際の曜日効果調整の基本となる時間領域での表現について述べる。第4節ではそれら二通りの表現の関係について検討する。第5節では調整をしようとする系列の曜日変動の構造が時間に沿って変化するという想定の下に確率的なモデルを設定し、実際のデータを用いて推定を行った結果についてまとめる。

2. 曜日周波数

ここでは日次で観察される系列を単純に月次集計することで作成される月次系列の性質について考察する。フローデータの曜日効果はデータの数値が曜日毎に異なっている系列が月次集計されることで生じる変動なので、その様な系列を形式的に構成し分析する。1週間の周期性を持つ日次データを月次集計した場合、集計データにも一定の周期的変動が現れる。それは原データのプロットから明かに見て取れる季節性の様な変動ではないが、周波数領域での分析においては曜日効果は特定の周波数“曜日周波数”におけるピークとして観測される。

曜日周波数についてはCleveland-Devlin [1] が詳細な分析を行っている。以下の記述の基本的な枠組みはCleveland-Devlin [1] に沿っているが、Cleveland-Devlin [1] は日次系列を連続的な過程と見て、月次集計を積分によって表現するという、やや不自然と思われる定式化を行っているため、ここでは離散的な日次系列の月次集計の形式について分析することで、Cleveland-Devlin [1] におけるaliasに関する議論等の煩雑な部分を回避し、(2.10)のような簡明な表現を導いた。また導出の過程はかなり異なっている。

まず $Y(d)$ を日次の系列とする。ここで d の単位は1日である。曜日変動を表現するために

$Y(d)$ が厳密に 7 日の周期を持つとする。このとき $Y(d)$ は三角関数を用いて次の様に表現することができる。

$$Y(d) = a_0 + \sum_{k=1}^3 \left\{ a_k \cos\left(\frac{2\pi kd}{7}\right) + b_k \sin\left(\frac{2\pi kd}{7}\right) \right\} \quad (2.1)$$

$a_0, a_k, b_k (k=1, 2, 3)$ はデータの 1 週間の変動パターンを規定するパラメーターである。次に t の単位を月として $Y(d)$ を月次に集計した系列を $y(t)^*$ とする。 d_t を観測されたデータの始めから t 番目にあたる月の最初の日を表すとすれば

$$\begin{aligned} y(t)^* &= \sum_{d=d_t}^{d_{t+1}-1} Y(d) \\ &= a_0 X_{0t}^* + \sum_{k=1}^3 (a_k X_{kt} + b_k Z_{kt}) \end{aligned} \quad (2.2)$$

と表現される。ここで X_{0t}^*, X_{kt}, Z_{kt} は以下の様に定義される。

$$X_{0t}^* = d_{t+1} - d_t \quad (2.3)$$

$$X_{kt} = \sum_{d=d_t}^{d_{t+1}-1} \cos\left(\frac{2\pi kd}{7}\right) \quad (2.4)$$

$$Z_{kt} = \sum_{d=d_t}^{d_{t+1}-1} \sin\left(\frac{2\pi kd}{7}\right) \quad (2.5)$$

X_{0t}^* は t 番目の月の日数を表している。またカレンダー上で曜日と日付が毎年 1 日ずつずれる事実と 4 年ごとの閏年を考慮すれば、 $y(t)^*$ は 7×4 年 = 336 ヶ月の周期を持っていることが分かる。¹

$X_{0t}^*, X_{kt}, Z_{kt} (k=1, 2, 3)$ の 7 つの変数は、実際のカレンダーから連続した 336 ヶ月を任意に選び d_t に当たる系列を指定すれば具体的に求める事ができる。

次に

$$\begin{aligned} X_{0t} &= X_{0t}^* - \frac{1}{336} \sum_{i=0}^{335} X_{0t+i}^* \\ &= X_{0t}^* - 30.4375 \end{aligned} \quad (2.6)$$

と置いて、

$$y(t) = a_0 X_{0t} + \sum_{k=1}^3 (a_k X_{kt} + b_k Z_{kt}) \quad (2.7)$$

を改めて曜日効果と定義する。ここで問題にするのは $y(t)^*$ の周期的特性なので、定数にあたる部分を $y(t)^*$ から除去した $y(t)$ の性質についてのみを見ることにする。 $y(t)$ は 336 ヶ月の周期で変動し、 $1/336 \sum_{i=1}^{336} X_{0i}^* = 30.4375$ であることから、 $y(t)$ の 1 周期 (336 ヶ月) での合計は 0 になる。従って (2.7) で定義される $y(t)$ は、曜日効果のみを表現する系列ということになる。なお定義より、 X_{0t} は月の長さが変化することによる効果を表しており、同時に *Leap Year effect* (閏年効果) も含んでいる。

また $X_{0t}, X_{kt}, Z_{kt} (k=1, 2, 3)$ の 7 つの系列はそれぞれ直交していることが確かめられるの

¹ 厳密には閏年の有無には 100 年単位の周期性があるので曜日と日付が完全に一致する周期は相当長いが、経済データの分析においては無意味なのでここでは無視することにする。

で、 $y(t)$ の二乗和は次式のように表すことができる。

$$\sum_{t=1}^{336} y(t)^2 = a_0^2 \sum_{t=1}^{336} X_{0t}^2 + \sum_{k=1}^3 \left(a_k^2 \sum_{t=1}^{336} X_{kt}^2 + b_k^2 \sum_{t=1}^{336} Z_{kt}^2 \right) \quad (2.8)$$

一方 $X_{0t}, X_{kt}, Z_{kt} (k=1, 2, 3)$ はいずれも 336 ヶ月の周期をもつ変数なので、これらに関して次のようにピリオドグラムを定義することができる。 $P_{Xk}(j)$ を X_{kt} に対応するピリオドグラムとすると

$$P_{Xk}(j) = \frac{1}{168} \left| \sum_{t=1}^{336} X_{kt} \exp\left(-i2\pi \frac{j}{336} t\right) \right|^2 \quad j=1, \dots, 168 \quad (2.9)$$

となる。また Z_{kt} に対応する $P_{Zk}(j)$ も同様に定義する。このとき $y(t)$ のピリオドグラムを $P_y(j)$ とおいて、 $P_{Xk}(j) = P_{Zk}(j) (k=1, 2, 3)$ となることを考慮すれば、 $j=1, \dots, 168$ に対して $P_y(j)$ は

$$\begin{aligned} P_y(j) &= \frac{1}{168} \left| \sum_{t=1}^{336} \left\{ a_0 X_{0t} + \sum_{k=1}^3 (a_k X_{kt} + b_k Z_{kt}) \right\} \exp\left(-i2\pi \frac{j}{336} t\right) \right|^2 \\ &= a_0^2 P_{X0}(j) + \sum_{k=1}^3 (a_k^2 + b_k^2) P_{Xk}(j) \end{aligned} \quad (2.10)$$

と書くことができるので、 $P_y(j)$ は $P_{Xk}(j) (k=0, \dots, 3)$ の線形結合により表現できることがわか

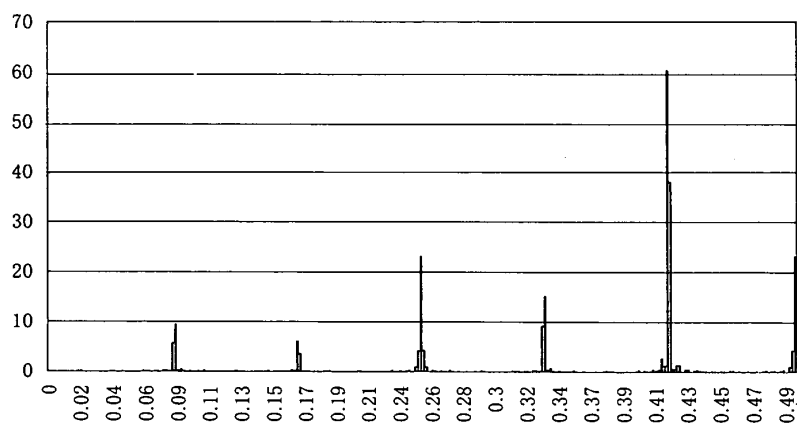


図1 $k=0$ のピリオドグラム

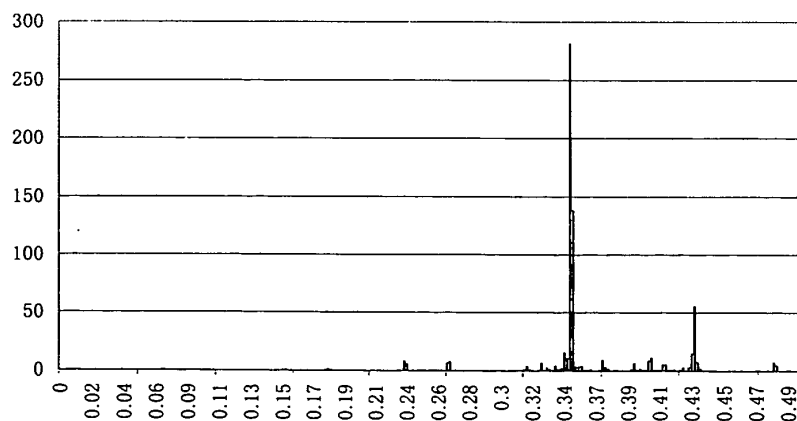


図2 $k=1$ のピリオドグラム

る。 $X_{0t}, X_{kt}, Z_{kt} (k=1, 2, 3)$ のとる値を実際に求め、(2.9)式に従い $P_{Xk}(j) (k=0, \dots, 3)$ を計算した。図1から図4は $P_{Xk}(j) (k=0, \dots, 3)$ を示したグラフである。以上より内部に7日の周期を含んでいる月次データは $P_{Xk}(j) (k=0, \dots, 3)$ によって表される周波数特性を持っていることがわかる。ただし7日周期の形状によって a_k と b_k の値は異なるため、それに応じて図1から図4のどのパターンが支配的になるかはデータにより異なってくる。 a_k と b_k は(2.7)式に従い、 $y(t)$ を $X_{0t}^*, X_{kt}, Z_{kt} (k=1, 2, 3)$ に回帰する事で推定できる。しかしながら $X_{0t}, X_{kt}, Z_{kt} (k=1, 2, 3)$ の二乗和を実際に計算した値は

$$\begin{aligned}\sum_{t=1}^{336} X_{0t}^2 &= 222.69 \\ \sum_{t=1}^{336} X_{1t}^2 &= \sum_{t=1}^{336} Z_{1t}^2 = 680.12 \\ \sum_{t=1}^{336} X_{2t}^2 &= \sum_{t=1}^{336} Z_{2t}^2 = 120.76 \\ \sum_{t=1}^{336} X_{3t}^2 &= \sum_{t=1}^{336} Z_{3t}^2 = 77.62\end{aligned}$$

となっており、これは a_k と b_k の値を評価する以前に、 $y(t)$ の変動に占める X_{kt}, Z_{kt} の割合にははじめからかなりの程度の違いがあることを示している。上の結果から、 X_{1t} および Z_{1t} の比率が特に高いので、そのピリオドグラムを示した図2において特に顕著な 0.348cycle/month およ

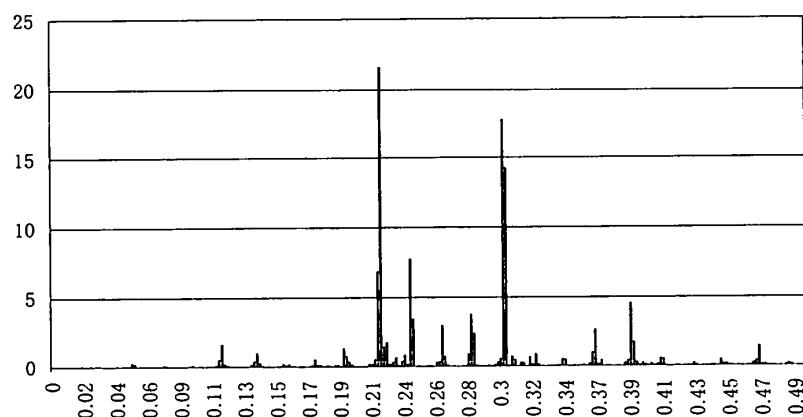


図3 $k=2$ のピリオドグラム

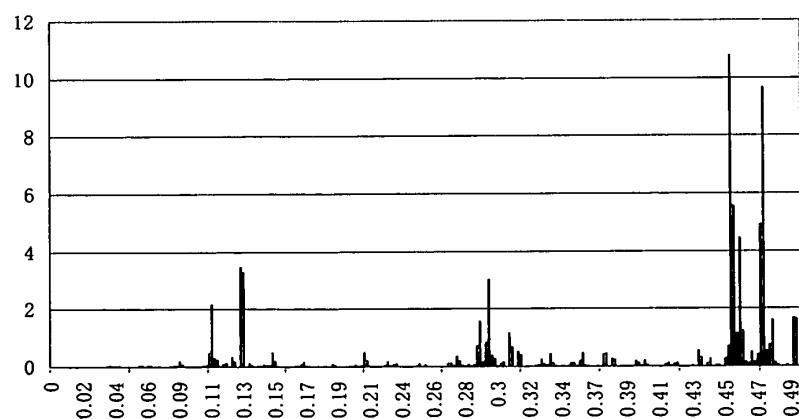


図4 $k=3$ のピリオドグラム

び0.432cycle/month付近でのピークを曜日効果の指標とすることが考えられる。よって分析対象のデータのスペクトルを推定した時にこれらの周波数で顕著なピークが見られるならば、データに曜日効果が存在する可能性が高いと言える。従ってスペクトルが推定されれば、 a_k と b_k に相当するパラメーターの推定を行わなくても、曜日効果の有無はある程度推測できることになる。

なお1ヶ月の長さの違いによる変動を意味する X_{0t} の周期は閏年を無視すれば1年であり、季節性と重なる。それはピリオドグラムを示した図1に見られる通り、 $1/12 \cong 0.082\text{cycle/month}$ とその整数倍のところにピークが存在することからも確認できる。従って適切な方法で季節調整された系列を問題にする限り、 X_{0t} の影響は考える必要性は少ないと思われる。

3. 時間領域における曜日効果の定義

曜日効果は、データの背後にある観測されない日次の系列に関して、各曜日がそれぞれ異なる水準を持つことから生じるので、月次のデータを各月に含まれるそれぞれの曜日の日数に回帰することにより推定するというのが、最も素朴な発想である。Dagun-Quennevill-Sutradhar [2]はそのアイディアを以下のように説明している。まず集計されたフローの月次時系列経済データは次の様に表現する事ができる。

$$y(t) = \sum_{j=1}^{N_t} y_j(t) \quad (3.1)$$

$y(t)$ はデータの第 t 月における値で、 $y_j(t)$ は第 t 月の j 番目の日の数値である。ここで N_t は第 t 月の日数を表す。これを変形して同一の曜日ごとにまとめると、

$$\begin{aligned} y(t) = & (y_1(t) + y_8(t) + \cdots + y_{29}(t)) \\ & + (y_2(t) + y_9(t) + \cdots + y_{30}(t)) \\ & + (y_3(t) + y_{10}(t) + \cdots + y_{31}(t)) \\ & \vdots \\ & + (y_7(t) + y_{14}(t) + \cdots + y_{28}(t)) \end{aligned} \quad (3.2)$$

ただしこれは $N_t=31$ となる月の例である。 i 曜日の平均値を $\bar{y}_i(t)$ ($i=1, 2, \dots, 7$)とし、 N_{it} を同月の i 曜日の日数とすると、

$$y(t) = \sum_{i=1}^7 N_{it} \bar{y}_i(t) \quad (3.3)$$

となる。さらに $\bar{y}_0(t) = 1/7 \sum_{i=1}^7 \bar{y}_i(t)$ とすれば

$$y(t) = \sum_{i=1}^6 (N_{it} - N_{7t})(\bar{y}_i(t) - \bar{y}_0(t)) + N_{7t} \bar{y}_0(t) \quad (3.4)$$

という形に変形できる。 $\bar{y}_0(t)$ はほぼ第 t 月の一日の平均の値を表していると考えられるので、(3.4)は $y(t)$ の変動を、曜日のずれによってもたらされる変動(右辺第1項)と1ヶ月の長さの変化によって生じる変動(右辺第2項)とに分解したものとみなすことができる。第1項はゼロの周りで周期的に変化する成分であり、第2項は第 t 月のレベルを表現している。

一般の月次経済時系列は季節性やその他の周期変動などを含んでいるのが普通であり、月次集計の元となる日次系列も季節的変動を有している。しかしここでは曜日効果のみを表現する

モデルを考えるために、 $y(t)$ が季節変動や周期変動を持たず、曜日のパターンも一定であるとの仮定を導入することにする。具体的には、 $y_i(t)$ が各 i について t によらず一定であるとする。すると $\bar{y}_i(t)$ と $\bar{y}_0(t)$ はいずれも t によらない定数になるので、(3.4) の右辺第 1 項はカレンダーの周期と同じ 336 ヶ月の周期を持っていることがわかる。第 2 項は第 t 月の日数に比例した平均レベルを表すことになる。従って閏年による変動 (leap year effect) も、この第 2 項に含まれる。

$\bar{y}_i(t)$ を時間 t に関して一定とみなした、固定係数の曜日効果モデルが、Young [6] によって提案されている。それは曜日のずれによる変動を表現する (3.4) の第 1 項だけを改めて曜日効果として定義し直して $z(t)$ とし、 $\bar{y}_i(t) - \bar{y}_0(t) = \eta_i$ と置いて

$$z(t) = \sum_{i=1}^6 (N_{it} - N_{7t}) \eta_i + e_t \quad (3.5)$$

なる線形回帰モデルの形式で曜日効果を表現するものである。ここで (3.4) の右辺第 2 項を除外する理由は、 $N_t \bar{y}_0(t)$ が $y(t)$ の平均的なレベルと一ヶ月の日数の変化に伴う季節的変動からなる成分であるために、トレンド + 季節性 + 曜日効果という時系列の分解において、トレンドや季節性と区別できなくなるからである。従って、この要素は季節調整やトレンドの抽出の過程の中で処理されることになる。またこのモデルにおいては、閏年による変動 (leap year effect) も無視されていることに注意する必要がある。

曜日効果の推定は、与えられた原系列から何らかの方法によりトレンド成分と季節成分を除去した後の系列を (3.5) における被説明変数 $z(t)$ として用いれば良い。なお $\bar{y}_i(t)$ がすべての i および t について同じ値を取るならば、明らかに 1 ヶ月に含まれる曜日の構成の変化による変動は存在しない。

4. 周波数領域での表現と時間領域での表現の関係

曜日効果の調整の基本的なアイデアは、前節で見たような形で、データを一ヶ月に含まれる月曜日から日曜日までの日数に回帰することである。つまり (3.5) のような形のモデルで説明される変動を取り除くことで得られる系列を曜日効果調整済系列とみなすわけである。

McNulty-Huffman [4] は月ごとの曜日の日数に基づいた、曜日効果の時間領域における表現に対応するピリオドグラムを導き、Cleveland-Devlin [1] が求めた曜日周波数におけるピリオドグラムのピークとの関係を議論している。

曜日効果を表す月次系列を $y(t)$ とし、次のように表記する。

$$y(t) = \sum_{i=1}^7 \gamma_i N_{it} \quad (4.1)$$

ここで N_{it} は t 月 i 曜日の日数で、 γ_i はその曜日の活動水準を表す。 $y(t)$ はすでに見たように 336 ヶ月の周期を持っている。ここで、 $N_{it} (i=1, 2, \dots, 7)$ という各曜日の日数を表す変数は、同一のものが時間的にずれて並んでいるだけであるということに注意すれば、この中の任意の一つを用いて書きかえることができる。1 年で曜日は 1 日ずれるが、閏年を考慮すると 4 年で 5 日ずれる。従って第 t 月の火曜日の日数を N_{Kt} とすれば、火曜日の 5 日前は木曜日なので、 N_{Kt} はその 48 ヶ月前の月の木曜日の日数に等しい。従って、 $N_{Kt} = N_{Kt-48 \times 2} = N_{Lt-48 \times 2} = N_{Mt-48 \times 3}$ という関係が成り立つ。第 t 月の水曜日の日数 N_{Kt} についても同様に考えれば、 $N_{Kt} = N_{Mt-48 \times 6}$ とすることができる。その他の曜日の日数も同じように考えて、過去の月曜日の日数で表現することを考えると、(4.1) は次の様に書き直すことができる。

$$y(t) = \gamma_1 N_{1t} + \gamma_2 N_{1t-48 \times 3} + \gamma_3 N_{1t-48 \times 6} + \gamma_4 N_{1t-48 \times 2} \\ + \gamma_5 N_{1t-48 \times 5} + \gamma_6 N_{1t-48} + \gamma_7 N_{1t-48 \times 4} \quad (4.2)$$

ただし N_{1t} を月曜日としている。項の順番を整えるために係数パラメーターを $\delta_0 = \gamma_1$, $\delta_1 = \gamma_6$, $\delta_2 = \gamma_4$, $\delta_3 = \gamma_2$, $\delta_4 = \gamma_7$, δ_5 , $\delta_6 = \gamma_3$ と置き直せば, (4.2) は

$$y(t) = \sum_{s=0}^6 \delta_s N_{1t-48s} \quad (4.3)$$

となる。(4.3) の意味するところは, $y(t)$ は $N_{it} (i=1, 2, \dots, 7)$ で表される 7 つの系列のうちのどれかひとつの系列に対してある線形フィルターをかけることで構成された系列であるとみなせるということである。従って y_t と N_{1t} のピリオドグラムをそれぞれ $P_y(j)$, $P_{N1}(j)$ とすれば, 次の様な関係が成り立っている。

$$P_y(j) = \left| \sum_{s=0}^6 \delta_s e^{-i2\pi s j / 7} \right|^2 P_{N1}(j) \quad j=1, \dots, 168 \quad (4.4)$$

$P_{N1}(j)$ を図 5 に示した。これは図 1 から図 4 で示された曜日効果の周波数領域での特徴を表す四つのパターンを合成した形になっており, いずれも同一の周波数においてピークを持っている事が分かる。つまり, 第 2 節で見た曜日効果に特徴的なピリオドグラムのいくつかのピークは, 基本的には曜日の日数の周期性に由来しており, 一週間の曜日のパターンがそのピークの高さを決定しているということになる。以上より (4.4) 右辺の前半の絶対値記号でくくられた部分は γ_i の値によって異なり, 第 2 節の設定での a_k と b_k の値に対応している。(4.1) で表される曜日効果を原データから差し引く操作が, 曜日周波数である 0.348 cycle/month 付近でのデータのスペクトルを削ることに対応していることがわかる。

5. 確率的係数モデルとその推定

5.1 曜日効果の固定係数モデルの問題点

第 3 節で示した固定係数モデル (3.5) は, 曜日効果の最も単純な表現であると言える。しかしながら曜日ごとの水準である η_i が分析対象とする時系列のデータ期間を通して一定である

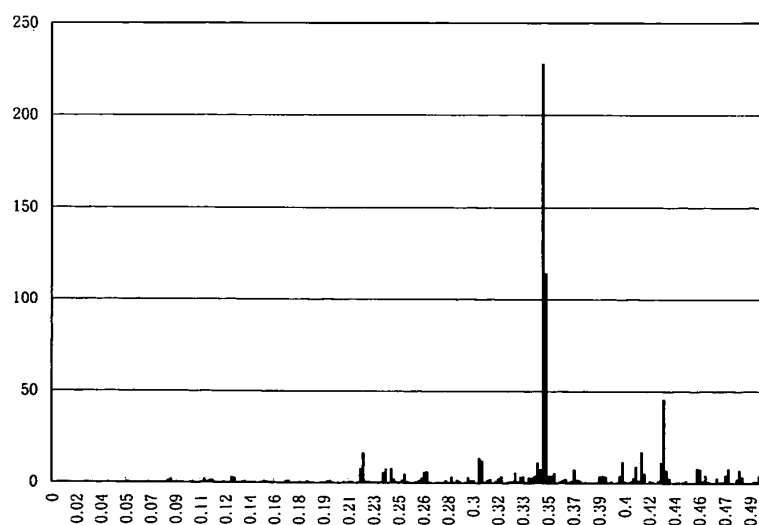


図 5 $P_{N1}(j)$ のグラフ

とする仮定が適当でない場合もあると考えられる。

図6は、1975年から1990年にかけての日本の鉱工業生産指数の原系列から、季節調整プログラム Decom (北川 [3]) を用いてトレンドサイクル成分と季節成分を推定し、残差の成分について推定したスペクトルをグラフに表したものである。時系列が、トレンド、季節性、不規則変動の3成分に分解できるという仮定の下では、図6は不規則変動のスペクトルを表すことになる。不規則な成分には通常はホワイトノイズを仮定するが、図6には曜日効果の存在を示唆する曜日周波数 0.348cycles/month 付近に明瞭なピークが見られる。従ってこの場合は曜日効果による変動を不規則変動の一部としてしまっていることになる。

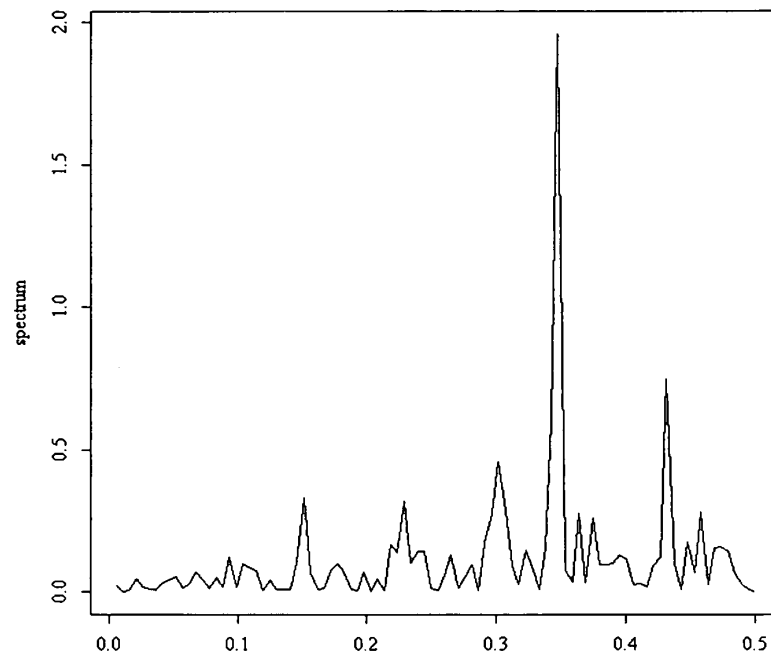


図6 鉱工業生産指数のスペクトル

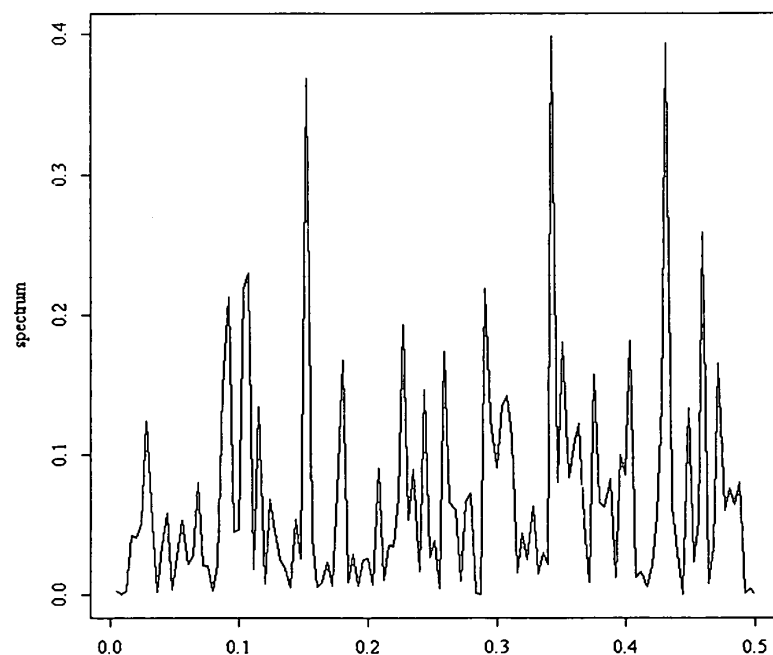


図7 固定係数モデルによって処理された鉱工業生産指数のスペクトル

図7は、同じデータに対して、以下に(5.3)として示した、(3.5)に若干の変更を加えた固定係数モデルをあてはめた場合の残差系列のスペクトルを推定したものである。これを見ると、まだ曜日周波数に相当する0.348cycles/monthおよび0.432cycles/month付近においてピークが残存しているため、固定係数モデルによっては曜日効果は表現し切れていない可能性があるといえる。

このようなことが起こる原因としては、曜日効果のパターンが変化していることが考えられる。第3節における表現においては曜日効果のパターンは a_k と b_k というパラメーターで特徴付けられていたが、すでに見たように a_k と b_k が多少変化しても0.348cycle/month付近でのピークは比較的明瞭に現れるため、曜日周波数のピークを見かけ上取り除くには、より観察できない曜日パターンの変化に即したモデルを用いることが必要である。

一つの対応としては、データを適当な区間に分割し、それぞれの区間で推定し直す方法などが考えられるが、この節では単純な線形回帰モデルである固定係数モデルの回帰係数が、時間 t に応じて確率的に変動することを許した確率的係数モデルを適用することを考え、その有効性について検討する。

5.2 曜日効果の確率的係数モデル

本節で提案するモデルは(3.5)を基本に、係数に確率的変動を導入したものである。ただし(3.5)では回帰モデルの説明変数を $N_{it}-N_{7t}$ という形で、一つの曜日の日数との差としているため、そのままでは回帰係数の持つ意味がややわかり難くなっているため、パラメーターの置き方に若干の変更を加えることにする。まず(3.4)式を次のように変形する。

$$y(t) = \sum_{i=1}^6 (N_{it} - N_{7t})(\bar{y}_i(t) - \bar{y}_0(t)) + (N_t - N_t^*)\bar{y}_0(t) + N_t^* \bar{y}_0(t) \quad (5.1)$$

N_t^* は

$$N_t^* = \frac{1}{4} \sum_{i=1}^4 N_{t+12i} \quad (5.2)$$

とする。定義より N_t^* は t 番目の月の日数を表し、2月にのみ28.25をとる変数となる。よって(5.1)式の右辺第1項及び第2項は曜日のずれによる変動と閏年効果を表現していると考えられる。第3項は一ヶ月の平均的なレベルと月の長さの違いによる変動を表現していると解釈できるので、これはトレンド成分と季節成分として処理されたと考えて、以後は第1項と第2項だけを改めて曜日効果と定義する。第3節での(3.4)に基づく曜日効果の定義とは異なり、閏年による効果を各曜日に分散して取りこんでいるということになる。従って(5.1)の右辺第1項及び第2項を用いて、まずは次のような固定係数の曜日効果モデルを考えることができる。

$$\begin{aligned} z(t) &= \sum_{i=1}^6 (N_{it} - N_{7t})(\bar{y}_i(t) - \bar{y}_0(t)) + (N_t - N_t^*)\bar{y}_0(t) + e_t \\ &= \sum_{i=1}^7 \left(N_{it} - \frac{1}{7} N_t^* \right) \bar{y}_i(t) + e_t \\ &= \sum_{i=1}^7 C_{it} \beta_i + e_t \end{aligned} \quad (5.3)$$

ただし一行目から二行目への変形で $\bar{y}_0(t) = 1/7 \sum_{i=1}^7 \bar{y}_i(t)$ と $N_t = \sum_{i=1}^7 N_{it}$ を用い、 $N_{it} - (1/7)N_t^* = C_{it}$ および $\bar{y}_i(t) = \beta_i$ と置いている。これにより(3.5)と較べて、閏年効果をモデルに組み込むとともに、月曜から日曜の活動水準を表す係数 $\bar{y}_i(t) = \beta_i$ が表面に現れて見通しを

良くすることができた。

次に (5.3) において、回帰係数に確率的変動を許す一つの方法として、次の 2 式で表現されるモデルを導入する。

$$y(t) = \sum_{i=1}^k C_{it}\beta_{it} + e_t = C_t'\beta_t + e_t \quad (5.4)$$

$$B(L)(\beta_t - \bar{\beta}) = \varepsilon_t \quad (5.5)$$

$$C_t = (C_{1t}, \dots, C_{kt})', \quad \beta_t = (\beta_{1t}, \dots, \beta_{kt})', \quad \bar{\beta} = (\bar{\beta}_1, \dots, \bar{\beta}_k)' \quad (5.6)$$

$$E[e_t] = 0, \quad \text{Var}[e_t] = \sigma^2, \quad E[\varepsilon_t] = 0, \quad \text{Var}[\varepsilon_t] = \Omega \quad (5.7)$$

C_t は説明変数ベクトル、 β_t は時間 t に伴って変動する回帰係数、 $\bar{\beta}$ は β_t の平均を表す。 e_t と ε_t は i.i.d. で互いに独立であるとする。また L はラグオペレーターで、 $B(L)$ は L の多項式である。 $B(L)$ にはモデルの設定に応じてパラメーターが含まれる。

よって (5.5) は β_t が多変量 AR 過程に従っていることを表しており、モデルの性質は $B(L)$ の形によって決まる。なお (5.4) に含まれるパラメーター k は、(5.3) に即せば $k=7$ であるが、一般的な表記を考え k とした。

5.3 推定の方法

(5.4) および (5.5) の組は状態空間表現として書きなおすことができるので、各誤差項に正規性を仮定する事により、予測誤差分解を用いて尤度関数が求められる。データが与えられればカルマンフィルタを利用して尤度が計算できるので、パラメータに関して数値的に最大化すれば (5.5) で表される確率過程が定常であるかどうかに関わらず各パラメータの最尤推定量を得る事ができる。これは時系列モデルにおける一般的な最尤法による推定である。

しかし (5.5) 式に定常性を仮定した場合には、より簡便な推定方法が考えられる。本節では定常性を仮定した上で、計算の容易な最小二乗法を繰り返す方法を用いた。

まず (5.4) を次の様に変形する。

$$\begin{aligned} y(t) &= C_t'\beta_t + e_t \\ &= C_t'\bar{\beta} + C_t'(\beta_t - \bar{\beta}) + e_t \\ &= C_t'\bar{\beta} + v_t \end{aligned} \quad (5.8)$$

ここで $C_t'(\beta_t - \bar{\beta}) + e_t = v_t$ と置いた。 $\bar{\beta}$ は定数ベクトルなので (5.8) 式は通常の線形回帰モデルと同じ形式をしているが、この場合の誤差項に相当する v_t は、系列相関と分散不均一性の両方を含む非定常な過程である。 $(\beta_t - \bar{\beta})$ が定常ならば v_t の分散、自己共分散が存在する。 $(\beta_t - \bar{\beta})$ の k 次自己共分散関数を

$$\Gamma_k = E[(\beta_t - \bar{\beta})(\beta_{t-k} - \bar{\beta})'] \quad (5.9)$$

とすると、次式が成り立つ。

$$\begin{aligned} E[v_t v_{t-k}] &= C_t'\Gamma_k C_{t-k} + \sigma^2 \delta_k \\ &= (C_{t-k} \otimes C_t)' \text{vec}(\Gamma_k) + \sigma^2 \delta_k \end{aligned} \quad (5.10)$$

$\otimes$ はクロネッカー積で、 $\text{vec}(A)$ は行列 A の列ベクトルを左から順に縦に並べたベクトルを表す。また δ_k は $k=0$ のとき 1、 $k \neq 0$ のとき 0 となる変数とする。

$W_{t,k} = C_{t-k} \otimes C_t$ ($k \neq 0$), $W_{t,k} = (1, (C_{t-k} \otimes C_t)')$ ($k=0$) および $\gamma_k = \text{vec}(\Gamma_k)$ ($k \neq 0$), $\gamma_k = (\sigma^2, \text{vec}(\Gamma_k)')$ ($k=0$) と置いて (5.10) に代入すると次の様な表現が得られる。

$$\begin{aligned} v_t v_{t-k} &= W'_{k,t} \gamma_k + (v_t v_{t-k} - E[v_t v_{t-k}]) \\ &= W'_{k,t} \gamma_k + v_{k,t}^* \end{aligned} \quad (5.11)$$

ここで $v_{k,t}^* = v_t v_{t-k} - E[v_t v_{t-k}]$ と置いた。 $v_{k,t}^*$ は平均ゼロの確率変数になる。

(5.8), (5.11) の2式から, $\bar{\beta}$ と γ_k の一つの推定量を次の様な2段階の手順で求める事が考えられる。

(1) 与えられたデータ $y(t) (t=1, \dots, T)$ を C_t に回帰し, $\bar{\beta}$ の推定量 $\hat{\beta}$ と, 残差 $\hat{v}_t = y(t) - C_t \hat{\beta}$ を得る。

(2) v_t は観測できないため, (5.8)式において v_t を $\hat{v}_t$ で置き換え, $\hat{v}_t \hat{v}_{t-k}$ を $W_{t,k}$ に回帰することにより γ_k の推定量 $\hat{\gamma}_k$ を得る。

ここで $\hat{\gamma}_0$ の第一要素は σ^2 の推定量 $\hat{\sigma}^2$ になっている。なお v_t を $\hat{v}_t$ に置き代えると, (5.11) より

$$\hat{v}_t \hat{v}_{t-k} = W'_{k,t} \gamma_k + v_{k,t}^* + (\hat{v}_t \hat{v}_{t-k} - v_t v_{t-k}) \quad (5.12)$$

と書けるので, 手順(2)での回帰は誤差項を $v_{k,t}^{**} = v_{k,t}^* + (\hat{v}_t \hat{v}_{t-k} - v_t v_{t-k})$ としたものになっている。 $\hat{v}_t \hat{v}_{t-k} - v_t v_{t-k}$ は推定量 $\hat{\gamma}_k$ の性質に影響を及ぼすが漸近的には無視し得る。

これらの推定量を $\mathbf{y} = (y(1), \dots, y(T))'$ および $\mathbf{C} = (C_1, \dots, C_T)'$ を用いて明示的に表すと次の2式になる。

$$\hat{\beta} = (\mathbf{C}'\mathbf{C})^{-1} \mathbf{C}'\mathbf{y} \quad (5.13)$$

$$\hat{\gamma}_k = \left(\sum_{t=1}^{T-k} W_{t,k} W'_{t,k} \right)^{-1} \sum_{t=1}^{T-k} W_{t,k} \hat{v}_t \hat{v}_{t-k} \quad (5.14)$$

このような形で得られる最小二乗法による推定量は一定の条件の下で一致性と漸近正規性を持つことが Pagan [5] によって示されている。従って $\hat{\beta}$ と $\hat{\gamma}_k$ については正規分布に基づいた検定を行うことができる。なお (5.13), (5.14) の推定量の標準誤差に関しては, 一般に相関構造が未知の誤差項を持つ回帰モデルの推定で用いられる, 次のような Bartlett のカーネルを用いた一致推定量 S_T を利用した。

$$S_T = M_T^{-1} V_T M_T^{-1} \quad (5.15)$$

$$M_T = \frac{1}{T} \sum_{t=1}^T \mathbf{x}_t \mathbf{x}_t' \quad (5.16)$$

$$V_T = \frac{1}{T} \left[\sum_{t=1}^T u_t^2 \mathbf{x}_t \mathbf{x}_t' + \sum_{i=1}^q \left[1 - \frac{i}{q+1} \right] \sum_{t=i+1}^T (\mathbf{x}_t u_t u_{t-i} \mathbf{x}_{t-i}' + \mathbf{x}_{t-i} u_{t-i} u_t \mathbf{x}_t') \right] \quad (5.17)$$

ここで $\mathbf{x}_t$ は回帰モデルの説明変数ベクトル, u_t は通常の最小二乗法によって得られた残差である。これを $\hat{\beta}$ および $\hat{\gamma}_k$ の推定の際に用いて, それぞれの標準誤差を求めた。

5.4 モデルの設定

実際に推定を行ったモデルは (5.5) が定常な VAR(1) に従うことを仮定したものである。VAR の次数についてはデータ数とパラメーター数の兼ね合いからあまり大きく取る事は難しい。具体的には以下の様になる。

$$y_t = C_t \beta_t + e_t \quad (5.18)$$

$$\beta_t - \bar{\beta} = \Phi(\beta_{t-1} - \bar{\beta}) + \varepsilon_t \quad (5.19)$$

Φ は 7×7 の母数行列, $\bar{\beta}$ は 7×1 の定数ベクトルである。その他の設定は (5.4), (5.5) と同じ

とする。推定したいパラメーターは e_t の分散 σ^2 と、 Φ 、および ε_t の共分散行列 Ω である。(5.14) から σ^2 と $\beta_t - \bar{\beta}$ の自己相関行列の推定値を得たら、ユール＝ウォーカー方程式と同様に Φ と Ω の各要素 $\Phi = \{\phi_{ij}\}$, $\Omega = \{\omega_{ij}\}$ に関する連立方程式を構成し、それを解くことにより Φ と Ω の推定値を求めれば良い。なおデータ数の制約から、 Φ と Ω が対角行列であるとする仮定をさらに付加した。従って推定したパラメーターは $\bar{\beta}$, $\phi_{i,i}$, $\omega_{i,i}$ ($i=1, \dots, 7$) および、 σ^2 である。このときの β_t の定常性の条件は、 Φ の各要素の絶対値が 1 より小さいことである。これは各曜日の水準の変動がそれぞれ独立であるとする仮定で、一般の VAR モデルから見ればかなりきつい制約を課することになる。しかしながら曜日効果とはそもそも月次データの背後に想定している観測されない日次データが、一週間の周期性をある程度安定して持っていることから生じる変動なので、曜日効果モデルはその一週間の周期性を緩やかに表現することがその目的となる。このモデルでの、月ごとの各曜日の平均的水準がそれぞれ一定の値の周りで若干の確率的変動をするという想定は、緩やかな周期性を表すためのものなので、月ごとの各曜日の平均的水準が月次データ上で互いに相関を持っているという仮定は、曜日効果の定義に照らせば、あまり意味があるとは考え難い。これが、パラメーター数を少なくするという以外に、 Φ と Ω を対角行列とすることの理由である。

5.5 結 果

前節に述べたモデルを実際のデータに適用し、パラメーターを推定した。使用したデータは 1975 年から 1993 年にかけての日本の鉱工業生産指数および売り場面積 1 平方メートルあたりの百貨店売上である。いずれのデータについても事前に季節調整プログラム Decomp (北川 [3]) を用いてトレンドサイクル成分と季節成分を除去し、残りの成分に対してモデルを適用した。まず第 1 段階の $\bar{\beta}$ の推定結果は表 1、表 2 の様になった。推定された係数をみると、鉱工業生産指数に関しては平日がほぼ同じ値で、日曜日が最も小さく、次いで土曜日が小さい値となった。百貨店の売上については、日曜日の値が大きく、その他の曜日はほぼ均等になった。これらの結果は生産と消費のデータに関しての直観に合致しているといえるだろう。ここまでは固定係数モデルを推定していることと同じなので、その効果を確認するためにもとの系列のスペクトルと残差のスペクトルをそれぞれ推定し比較した。鉱工業生産指数についてはすでに見たが、百貨店売上の結果について、原系列のスペクトルを図 8 に、調整後のスペクトルを図 9 にそれぞれ示した。これによれば調整前には顕著に見られた 0.348 付近でのピークが固定係数モデルによって十分に取り除かれていると判断できる。なお図 9 では 0.1 付近でのピークの残存が見られるが、これは曜日周波数には相当しないため、曜日効果以外のこのデータに固有の性質と考えられる。なお残差に前節で定義した確率的モデルを適用してみたが、いずれの母数も有意には推定されなかった。

鉱工業生産指数に確率的係数モデルを適用した結果は表 3 の通りである。すべての曜日に

表 1 鉱工業生産指数

	推定値	標準誤差	t 値
日曜	0.546095	0.108614	5.027854
月曜	0.956654	0.104394	9.163881
火曜	0.927953	0.109400	8.482186
水曜	0.782189	0.102653	7.619713
木曜	0.969469	0.103406	9.375321
金曜	1.013658	0.107528	9.426951
土曜	0.730543	0.110920	6.586197

表 2 百貨店売上

	推定値	標準誤差	t 値
日曜	0.121015	0.025406	4.763282
月曜	0.090165	0.027395	3.291288
火曜	0.066949	0.026775	2.500380
水曜	0.094004	0.026769	3.511667
木曜	0.078843	0.027420	2.875359
金曜	0.078816	0.025403	3.102614
土曜	0.073678	0.028290	2.604433

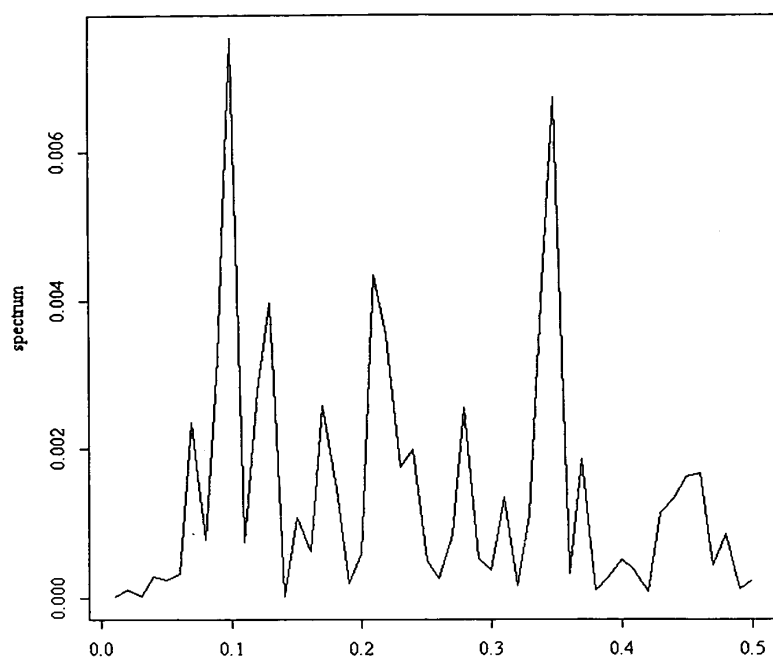


図8 百貨店売上のスペクトル

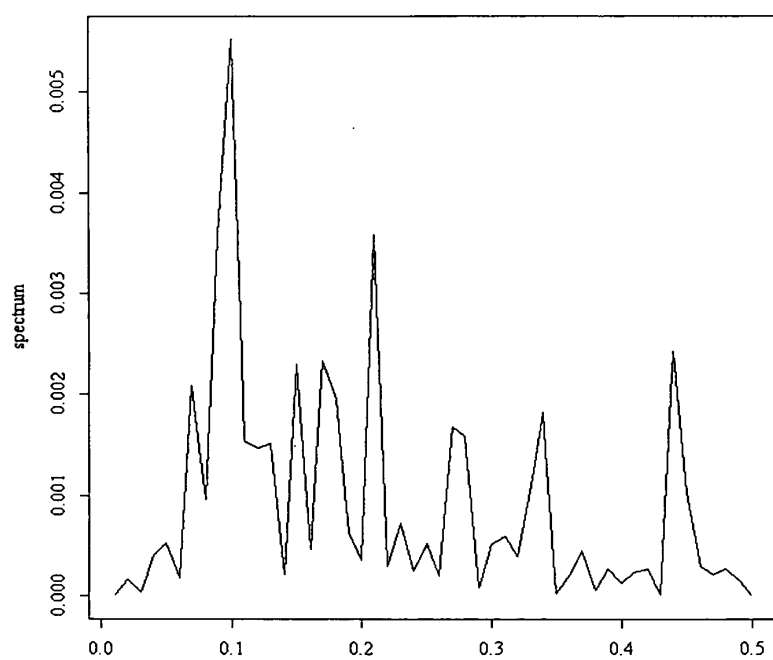


図9 固定係数モデルによって処理された百貨店売上のスペクトル

表3 鉱工業生産指数

	$\phi_{i,i}$ の推定値	$\omega_{i,i}$ の推定値	σ^2 の推定値
日曜	0.142292	0.368987	0.00062
月曜	0.022247	0.050008	
火曜	-0.03379	0.03395	
水曜	0.064229	0.148549	
木曜	-0.08344	0.02602	
金曜	0.011502	0.150457	
土曜	-0.06505	0.02962	

いて回帰を行ったところ日曜日の変動に関わるパラメーター以外については $\phi_{i,i}$ は、通常の有意水準では 0 であるとする帰無仮説を棄却できなかった。しかし日曜日とその他の曜日との間で推定値に関してに一定の差異が見られるため、説明変数を日曜日、土曜日、平日という形にまとめて足し合わせることで新たな 3 つの説明変数を作成し、 Φ と Ω を 3×3 の行列として推定を行ったところ、表 4 の様な結果を得た。これは各曜日の係数のうち、固定的な部分に関しては通常の回帰により求め、確率的な変動については日曜、土曜、平日の 3 種類で推定したことになる。この結果を用いてデータを調整し、スペクトルを推定したところ図 10 のようになった。図 7 と図 10 の比較からは、この確率的係数モデルによる調整によって、全体のレベルがやや下がるとともに、曜日周波数である 0.348 付近と 0.432 付近でのピークが特に際立って削られている様子が見てとれる。ただし 0.225 付近のごく狭い範囲でのみ調整後のレベルが上がっている個所があり、これについては定常な VAR(1) モデル以外も考慮したより適切なモデル選択の必要性を示唆していると言えるかもしれない。なお、図 11 から図 13 は表 4 に示したパラメーターの推定結果を所与とした上でカルマンフィルターによる状態推定を行い、確率的係数

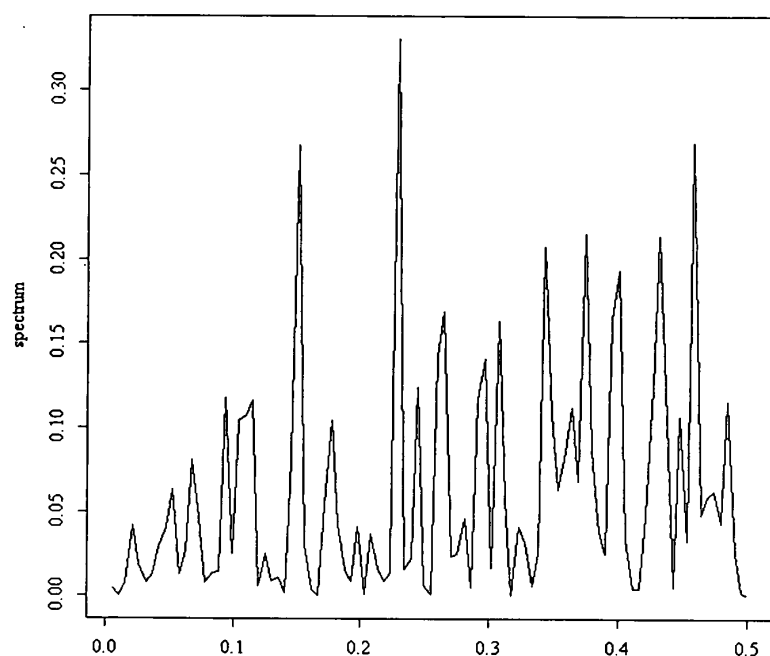


図 10 確率的係数モデルによって処理された鋳工業生産指数のスペクトル

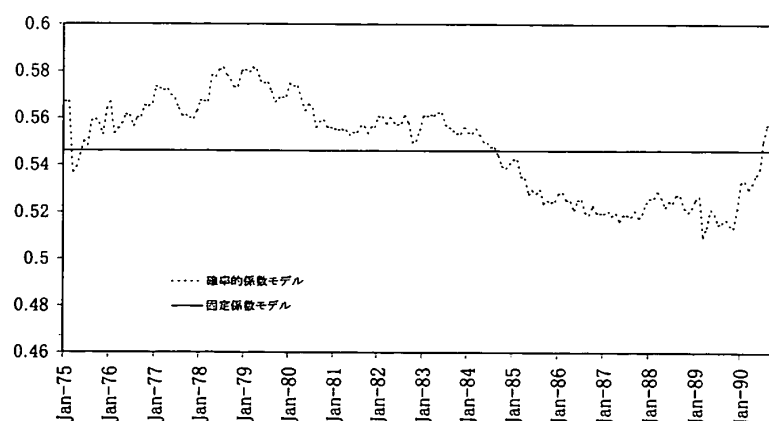


図 11 日曜日の係数の変動

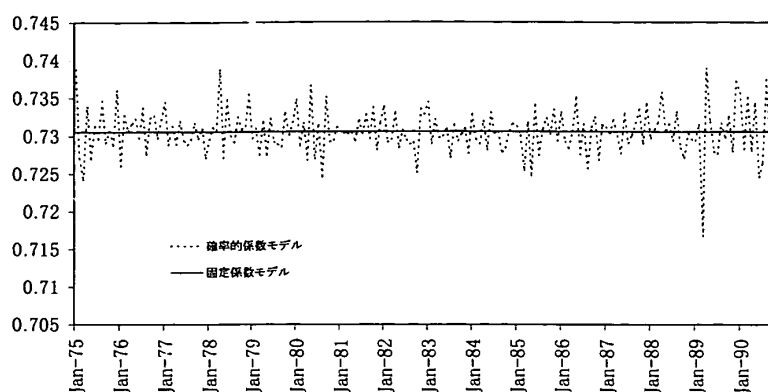


図12 土曜日の係数の変動

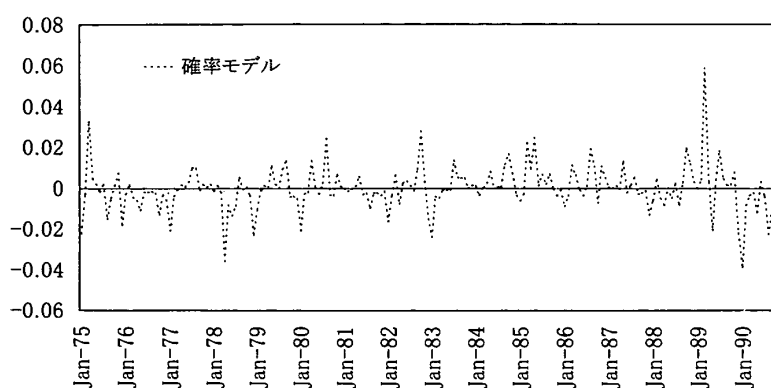


図13 平日の係数の変動

表4 鉱工業生産指数

	$\phi_{i,i}$ の推定値	$\omega_{i,i}$ の推定値	σ^2 の推定値
日曜	0.869256	0.042835	0.06125
平日	0.307776	0.049271	
土曜	-0.44339	0.022972	

の時間に沿った変化をグラフに表したものである。これらの図からは、各曜日の変動がかなり異なる統計的性質を有している可能性がうかがわれる。

6. ま と め

曜日効果モデルの定義とその周波数特性について概観し、固定係数モデルを拡張した確率的係数モデルを実際のデータに適用した。その結果、スペクトルの比較という観点からは、固定係数モデルに対して一定の改善が見られた。すなわち曜日周波数においてスペクトルにピークを持つ過程を、確率的係数モデルがより柔軟に表現していた。

曜日効果に関するその他の研究としては Dagun-Quennevill-Sutradhar [2] が (3.5) の表現を基に、 η_i がランダムウォークに従うと仮定したモデルを百貨店の売上データに適用し受け入れている。そのようなアプローチは、推定すべきパラメータを節約できるという実用上のメリットはあるが、本論の結果からは固定係数モデルとランダムウォークを利用した非定常モデルの中間にあたるケースとして、 β_i が定常過程に従うという仮定が妥当する場合もあると思われる。また Dagun-Quennevill-Sutradhar [2] の想定は η_i が VAR(1) に従うが全ての i について

ARの係数が1であるという仮定と同じなので、より強い制約を課していると言える。本論で考察した β_t が定常過程に従うと仮定したモデルは、係数が固定の場合もランダムウォークとなる場合も特別なケースとして含んでおり、また各曜日が異なるパラメーターを含んだ別個の確率モデルに従うことを許すという点でも、より一般的な形式になっている。データによっては定常VARが不適切なケースは当然ありうるが、推定されたモデルを用いて予測をするという観点からは、可能なならば定常なモデルの方が望ましいであろう。

今後の課題のひとつとしては $B(L)$ の形をデータからどのように決定するかという問題がある。特にモデルの特性方程式が単位根を持つ場合は今回用いた最小二乗法による推定量は望ましい性質を持たない。VARモデルを適用する場合にも、可能なパラメーターの組み合わせが多くなるため、実用上はAIC等を用いた客観的基準による機械的なモデルの選択ができた方がよい。ただし推定すべきパラメーターの個数が多い場合の最尤推定は、尤度関数の数値的最大化が難しいという、実際の計算上の問題が生じる可能性がある。この点は最小二乗推定量を用いることにメリットがあると言えるが、最小二乗推定量を初期値として与えて最大化を行う方法が考えられる。また本論での曜日効果の推定には、トレンドと季節性を除いた系列を用いたため、トレンドの抽出を目的としたデータ解析においては、曜日効果はあまり問題とはならないが、トレンドと季節性を表現するモデルと曜日効果モデルを合わせて、同時に未知パラメーターを最尤法などにより推定する場合には、曜日効果モデルの表現がトレンドの推定にも影響する。

一方、推定されたスペクトルに基づく視覚的な曜日調整法の評価は、曜日効果の時間領域における真のモデルが正確に同定されたかどうかには依存しないという点で有用であると言えるが、評価に客観性を持たせるためには何らかの基準が必要と思われる。

またその他の可能性としては、今回の確率的モデルにおけるパラメーターの変動の推定結果から推測する限り、その他の考えられるモデルとして、ランダムウォークに従う係数と定常過程に従う係数が混在しているようなモデルなどがあり得るかもしれない。例えば消費に関するデータなどでは、平日に関しては水準に大きな変動はないが、土曜日や日曜日の水準については時間を追って大きく変化しているという想定にも一定の合理性があると思われる。今回用いたモデルを利用して、各曜日の変動の傾向を別個に推定することで、曜日ごとの変動の特徴など、データの性質の分析に資するような知見が得られる可能性がある。これらの点を踏まえてさらにデータ分析が必要であると言える。

参 考 文 献

- [1] Cleveland, W. S. and Devlin, S. J. (1980). Calendar Effects in Monthly Time Series; Detection by Spectrum Analysis and Graphical Methods, *Journal of American Statistical Association*, 75, 487-496.
- [2] Dagum, E. B., Quenneville, B. and Sutradhar, B. (1992). Trading day Variation Multiple Regression Models with Random Parameters, *International Statistical Review*, 60, 57-73.
- [3] 北川源四郎 (1993). 時系列解析プログラミング, 岩波書店.
- [4] McNulty, M. S. and Huffman, W. E. (1989). The Sample Spectrum of Time Series with Trading Day Variation, *Economics Letters*, 31, 367-370.
- [5] Pagan, A. R. (1980). Some identification and estimation results for regression models with stochastically varying coefficients, *J. Econometrics*, 13, 341-363.
- [6] Young, A. H. (1965). Estimating Trading-Day Variation in Monthly Economic Time Series, *Technical Paper 12, Bureau of the Census, U.S. Department of Commerce*.