

第3章

商品分類の改訂に伴う貿易統計の整合性評価

野田容助

はじめに

本章では世界貿易統計データを利用するにあたって商品分類の改訂によって生じる時系列的接続の不整合の評価方法を紹介し、また、この方法をOECD貿易統計における日本のデータに適応した結果を検討することを目的としている。日本のデータでは対象年度を1962年から99年度までとして、相手国は世界計を使用する^(註1)。評価の方法としては分類コードおよび商品グループに対応する貿易取引額の時系列データが商品分類の改訂前後で断層が生じているかどうかを検討の基準とする。

貿易統計の時系列データはその国の経済変動および貿易の取引状況により年ごとに大きく変化する傾向があり、また、年データを利用しているという制約条件のために、商品分類の改訂の影響がないときでさえも必ずしも滑らかな変動をすることは限らない。しかし、断層の存在が検証できればその断層の要因の1つは商品分類の改訂によって生じた分類コードおよび商品グループの対応関係の不整合による可能性があると考えられるわけである。

一般的に時系列データに構造変化が生じているときにその時点を変化点という。変化点が商品分類の改訂時点に一致する分類コードおよび商品グループは必ずしも不整合の存在を意味しているわけではないが、不整合であれば変化点

が商品分類の改訂時点に一致するという状態はおこる可能性がある。商品分類の改訂時点と変化点が一致していることは、商品分類の変換にともなう不整合性の評価に対する必要条件とみなすことができる。このように改訂時点を境として生ずる取引額の変化を時系列の構造変化と見なし、本章ではこの構造変化に対してデータを商品分類の改定期時点近くで分割する回帰モデルと変化点をパラメータとしたときのその後分布という2つの方法を適用して判定することにする。

この2つの方法において分類コードおよび商品グループとして示された取引額の時系列が商品分類の改訂の時点、すなわち、SITC-R1からSITC-R2の改訂時点である1978年と、SITC-R2からSITC-R3の改訂時点である1988年の2時点を境として大きく変化していれば、その原因は商品分類の改訂にともなう対応関係の不整合によって生じたか、あるいは対応関係とは無関係な貿易統計自身に現れた経済変動によるものと考えられる^(註2)。

1. 構造変化の評価方法

時系列の変化点を調べるための統計的な方法としてはStepwise Chewテスト、スイッチング回帰モデル (Switching regression model) や赤池情報量基準AIC (Akaike Information Criterion) に

よる線形回帰モデルのモデル選択、ベイズにもとづく事後分布等の方法がよく利用される。一般的には変化点の推計はその個数が知られていないことが多く、変化点を見つけるに先立ってその個数を推計することが重要な課題となる。

しかし、本章では商品分類の改訂のあった年を境として商品グループの取引額の時系列が対応関係の不整合により大きく変化したかどうかを検討することを目的としている。したがってその取引額の推移がSITC-R1からSITC-R2へと改訂された1978年とSITC-R2からSITC-R3へと改訂された1988年に変化したかどうかを知りたいので、変化点の個数は2回と仮定することができる。変化点が商品分類の改訂時点と一致していることがわかれば、それは商品分類の改訂時点において構造変化が生じたと見なすことができるわけであるから、その商品グループは商品分類の改訂にともなう対応関係の接続に検討すべき内容がある可能性を示していると考えられる。

1.1 Box-Cox変換

時系列データは平均値の増加にともなう分散も増加したり、分布が著しく偏ってくることが多い。このようなデータは分布が正規分布に従って分布することを前提とする統計解析においてはうまくデータを処理できない。こうしたデータに対して変動がほぼ一定の状態を保つようなデータを得るための一般的なデータ変換にBox-Cox変換がある。時系列 $\{z_1 \dots z_n\}$ と $\{y_1 \dots y_n\}$ があるとき、

$$z = \begin{bmatrix} z_1 \\ \vdots \\ z_n \end{bmatrix} \quad y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$$

と表すことにする。 y から z への変換は、

$$(1) \quad z = \begin{bmatrix} h_1(y) \\ \vdots \\ h_n(y) \end{bmatrix}$$

と表すことができる。Box-Cox変換は(1)式において、

$$(2) \quad h_j(y) = \begin{cases} \frac{y_j^\lambda - 1}{\lambda} & \lambda \neq 0 \\ \log(y_j) & \lambda = 0 \end{cases}$$

$$(j=1 \dots n)$$

となる変数変換である^(註3)。ここで、 λ はパラメーターである。Box-Cox変換は定数を見れば、 $\lambda=0$ のときは対数、 $\lambda=1$ は逆数、 $\lambda=0.5$ は平方根、 $\lambda=1$ は原データと一致する。データに適した変換のパラメーター λ はAICの最小化によって選択することができる。

確率変数 y の同時密度関数を $g(y_1 \dots y_n)$ 、 z のそれを $f(z_1 \dots z_n)$ とする。 y から z への変数変換をおこなうとこれらの密度関数の関係は、

$$(3) \quad g(y_1 \dots y_n) = f(z_1 \dots z_n) |J|$$

と表される。ここで、 J はヤコビアン、 $|J|$ は絶対値を表し、(1)式と(2)式より、

$$J = \frac{\partial(z_1 \dots z_n)}{\partial(y_1 \dots y_n)}$$

$$= \begin{bmatrix} y_1^{\lambda-1} & & \\ & \ddots & \\ & & y_n^{\lambda-1} \end{bmatrix}$$

である。(3)式において $f(z_1 \dots z_n)$ に代わって平均 μ と分散 σ^2 を持ち、 z に関して独立な正規分布 $\phi(z_1 \dots z_n)$ をあてはめたときの尤度関数を考える。

$$L_g(\mu, \sigma^2 | y_1 \dots y_n) = L_\phi(\mu, \sigma^2 | z_1 \dots z_n) |J|$$

$$= (2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2} \sum_{j=1}^n (z_j - \mu)^2} |J|$$

ここで、

$$L_g(\mu, \sigma^2 | y) = g(y) \quad L_\phi(\mu, \sigma^2 | z) = \phi(z)$$

とする。(2)式よりこの尤度関数は λ の関数である。この尤度関数の対数をとれば、

$$\begin{aligned} \ell_g(\theta, \sigma^2 | y) &= \ell_\phi(\mu, \sigma^2 | z) + \log(|J|) \\ (4) \quad &= -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{j=1}^n (z_j - \mu)^2 \\ &\quad + \log(|J|) \end{aligned}$$

と表わすことができ、対数尤度関数という。この尤度関数も λ の関数である。

この対数尤度関数を μ と σ^2 により偏微分したものを0とおいた連立方程式を正規連立方程式という。

$$\frac{\partial \ell_g(\mu, \sigma^2 | z)}{\partial \mu} = 0 \quad \frac{\partial \ell_g(\mu, \sigma^2 | z)}{\partial \sigma^2} = 0$$

この正規連立方程式を解いて得られる解の $\hat{\mu}$ と $\hat{\sigma}^2$ を最尤推定量といい、

$$\hat{\mu} = \frac{1}{n} \sum_{j=1}^n z_j \quad \hat{\sigma}^2 = \frac{1}{n} \sum_{j=1}^n (z_j - \hat{\mu})^2$$

と計算される。これらの結果を(4)式の対数尤度関数に代入すると最大対数尤度が得られ、

$$\begin{aligned} (5) \quad \hat{\ell}_g(\hat{\mu}, \hat{\sigma}^2 | y) \\ = -\frac{n}{2} \{\log(2\pi\hat{\sigma}^2) + 1\} + \log(|J|) \end{aligned}$$

となる。

対数尤度関数(4)式の自由なパラメーターの数は平均 μ と分散 σ^2 の2個であり、(5)で得られた最大対数尤度は λ の関数である。したがって、(4)式のAICは λ の関数として表され、

$$\begin{aligned} \text{AIC}_g(\lambda) &= -2(\text{最大対数尤度}) \\ &\quad + 2(\text{自由なパラメーター数}) \\ &= n\{\log(2\pi\hat{\sigma}^2) + 1\} - 2\log(|J|) + 4 \end{aligned}$$

が得られる。AIC_g(λ)が最小になる λ がBox-Cox変換の最適な λ となる。

1.2 回帰モデルによる変化点の推定

時系列データ $\{y_1 \dots y_n\}$ を貿易の取引額とする。SITC-R1からSITC-R2への改訂に関しては時系の存在するすべての範囲のデータを使用せずに、改訂に影響する1962年から1987年までの年別データのみを使用し、この区間のデータを2

つに分割する。この分割は、SITC-R1の適用期間を $[62, m]$ 、SITC-R2の適用期間を $[m+1, 87]$ と仮定することである。分割されたそれぞれの期間において次のように線形回帰式の当てはめをおこなう。

パラメーターを $\{\alpha_i, \beta_i\}$ $i=1, 2$ とすると、線形回帰式は次のように分割して表わされる。

$$(6) \quad y_i = \begin{cases} \alpha_1 + \beta_1 x_i + e_i & i=1, \dots, m \\ \alpha_2 + \beta_2 x_i + e_i & i=m+1, \dots, n \end{cases}$$

ここで、攪乱項の e_i は正規分布に従っていると仮定する。

$$e_i \sim N(0, \sigma^2) \quad i=1 \dots n$$

分割の時点である m は1963年から85年までの範囲を動くように定める。

SITC-R2からSITC-R3への改訂に関しては、改訂に影響する1978年からデータの存在する1999年までのデータのみを使用し、この区間のデータを2つに分割する。この分割は、SITC-R2の適用期間を $[78, m]$ 、SITC-R3の適用期間を $[m+1, 93]$ と仮定することである。分割されたそれぞれの期間において(6)式のように線形回帰式の当てはめをおこす。分割時点 m の動く範囲は1979年から97年までとする。

区間を分割して表された線形回帰式の(6)式を行列で表現すると、

$$(7) \quad y = X\theta + e$$

となる。ここで、 y は観測値の n 次ベクトル e は攪乱項の n 次ベクトルであり、

$$y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \quad e = \begin{bmatrix} e_1 \\ \vdots \\ e_n \end{bmatrix}$$

とする。 e は、

$$(8) \quad e \sim N(0, \sigma^2 I)$$

であり、その同時密度関数は、

$$(9) \quad f(e) = (2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2} e'e}$$

と表される。また、 X は $n \times 4$ のマトリクス、 θ は4次のベクトルとする。 X は分割に時点 m によ

って変化する。

$$X = \begin{bmatrix} 1 & x_1 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_m & 0 & 0 \\ 0 & 0 & 1 & x_{m+1} \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 1 & x_n \end{bmatrix} \quad \theta = \begin{bmatrix} \alpha_1 \\ \beta_1 \\ \alpha_2 \\ \beta_2 \end{bmatrix}$$

このとき (7) のモデルの尤度関数は (9) より、

$$L(\theta, \sigma^2 | y_1 \cdots y_n) = f(y | \theta, \sigma^2) \\ = (2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2}(y-X\theta)'(y-X\theta)}$$

である。この対数尤度関数をとれば、

$$\begin{aligned} \ell(\theta, \sigma^2 | y) &= \log L(\theta, \sigma^2 | y) \\ (10) \quad &= -\frac{n}{2} \log(2\pi\sigma^2) - \frac{(y-X\theta)'(y-X\theta)}{2\sigma^2} \end{aligned}$$

と表される。 θ と σ^2 の最尤推定量は対数尤度関数を θ と σ^2 により偏微分したものを 0 とおいた正規連立方程式から得られる。

$$\frac{\partial \ell(\theta, \sigma^2 | y)}{\partial \theta} = 0 \quad \frac{\partial \ell(\theta, \sigma^2 | y)}{\partial \sigma^2} = 0$$

その解は、

$$\begin{aligned} \hat{\theta} &= (X'X)^{-1} X'y \\ (11) \quad \hat{\sigma}^2 &= \frac{(y-\hat{y})'(y-\hat{y})}{n} \end{aligned}$$

となる。これらの結果を対数尤度関数に代入すると最大対数尤度が得られ、

$$\hat{\ell}(\hat{\theta}, \hat{\sigma}^2 | y) = -\frac{n}{2} \{\log(2\pi\hat{\sigma}^2) + 1\}$$

となる。ここで、 \hat{y} は y の推定量であり、(11) で得られた最尤推定量 $\hat{\theta}$ により、

$$(12) \quad \hat{y} = X\hat{\theta}$$

である。このモデルの自由なパラメーターの数は θ の要素の個数である 4 個と分散 σ^2 を合わせて 5 個である。 X は分割の時点 m によって変化するので、最大対数尤度は m の関数である。したがって、(7) 式のモデルの AIC は m の関数として表され、

$$(13) \quad \text{AIC}(m) = n\{\log(2\pi\hat{\sigma}^2) + 1\} + 10$$

与えられる。最小となる $\text{AIC}(m)$ を持つ m を変化点の推計値とする。

1.3 事後分布による変化点の推定

線形回帰式の (7) 式において、パラメーターとして使用されているのは θ, m, σ^2 であるが、(7) 式では未知の固定された値として定義されている。それに対して、ここではこれらのパラメーターは確率変数として定義することにする。

確率変数 m, θ, σ^2 に関する事前分布の密度関数をそれぞれ、

$$f(m) = \begin{cases} \frac{1}{n-3} & m=2, \dots, n-2 \\ 0 & \text{otherwise} \end{cases}$$

$$f(\theta) \propto c$$

$$f(\sigma^2) \propto \frac{1}{\sigma^2}$$

とする。ここで、 α は比例関係を表わし、 c は定数とする。データ y が得られたとき、(10) 式を利用すれば確率変数 m, θ, σ^2 に関する事後分布の同時密度関数はベイズの定理により、

$$\begin{aligned} f(m, \theta, \sigma^2 | y) \\ (14) \quad &= \frac{f(m, \theta, \sigma^2) f(y | m, \theta, \sigma^2)}{f(y)} \\ &= c_2 [\sigma^2]^{-\frac{n+2}{2}} e^{-\frac{1}{2\sigma^2}(y-X\theta)'(y-X\theta)} \end{aligned}$$

となる^(註4)。 c_2 は定数とする。 m と σ^2 に関する周辺分布を求めるため (12) 式の同時密度関数を θ について積分する準備をする。

$$\begin{aligned} &(y-X\theta)'(y-X\theta) \\ &= [(y-\hat{y}) - X(\hat{\theta}-\theta)]' [(y-\hat{y}) - X(\hat{\theta}-\theta)] \\ &= (y-\hat{y})'(y-\hat{y}) + (\hat{\theta}-\theta)' X' X (\hat{\theta}-\theta) \end{aligned}$$

これは、

$$\begin{aligned} (y-\hat{y})X &= [y - X(X'X)^{-1}X'y]' X \\ &= y'[X - X] = 0 \end{aligned}$$

となることから得られる。したがって、 m と σ^2

に関する周辺密度関数は、

$$f(m, \sigma^2 | y) = c_2 [\sigma^2]^{\frac{n+2}{2}} e^{-\frac{1}{2\sigma^2}(y-\hat{y})'(y-\hat{y})} \int e^{-\frac{1}{2\sigma^2}(\hat{\theta}-\theta)'X'X(\hat{\theta}-\theta)} d\theta$$

で得られる。(7)式より、

$$X'y = X'X\theta + X'e$$

であり、 $X'X$ は 4×4 の正則行列なので逆行列が存在する。さらに、 $X'e = \xi$ とすると、 ξ は θ と同じ4次元ベクトルである。

$$\theta = (X'X)^{-1}(X'y + \xi)$$

このヤコビアン行列は、

$$\left| \frac{\partial \theta(\xi)}{\partial \xi} \right| = |(X'X)^{-1}|$$

である。これを利用して、

$$\begin{aligned} & \int e^{-\frac{1}{2\sigma^2}(\hat{\theta}-\theta)'X'X(\hat{\theta}-\theta)} d\theta \\ &= \int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2}\xi'(X'X)^{-1}\xi} |(X'X)^{-1}| d\xi \\ &= |X'X|^{-\frac{1}{2}} (2\pi\sigma^2)^2 \end{aligned}$$

となる^(註5)。この結果として、

$$\begin{aligned} f(m, \sigma^2 | y) &= c_2 [\sigma^2]^{\frac{n+2}{2}} |(X'X)^{-1}| \sigma^2 \frac{1}{2} e^{-\frac{1}{2\sigma^2}(y-\hat{y})'(y-\hat{y})} \\ &= c_2 |X'X|^{-\frac{1}{2}} [\sigma^2]^{\frac{n-2}{2}} e^{-\frac{a}{\sigma^2}} \end{aligned}$$

である。ここで、

$$a = \frac{(y-\hat{y})'(y-\hat{y})}{2}$$

としており、 c_2 は定数である。

最後に、 m に関する周辺分布を得るために σ^2 について積分する。 Γ 関数を利用すれば、

$$\begin{aligned} & \int_0^{\infty} [\sigma^2]^{\frac{n-2}{2}} e^{-\frac{a}{\sigma^2}} d\sigma^2 \\ &= a^{-\frac{n-4}{2}} \Gamma\left(\frac{n-2}{2}\right) \end{aligned}$$

となる^(註6)。これらの結果から、 m に関する事後分布の周辺密度関数は、

$$(15) \quad f(m|y) = c_3 |X'X|^{-\frac{1}{2}} \{(y-\hat{y})'(y-\hat{y})\}^{-\frac{n-4}{2}}$$

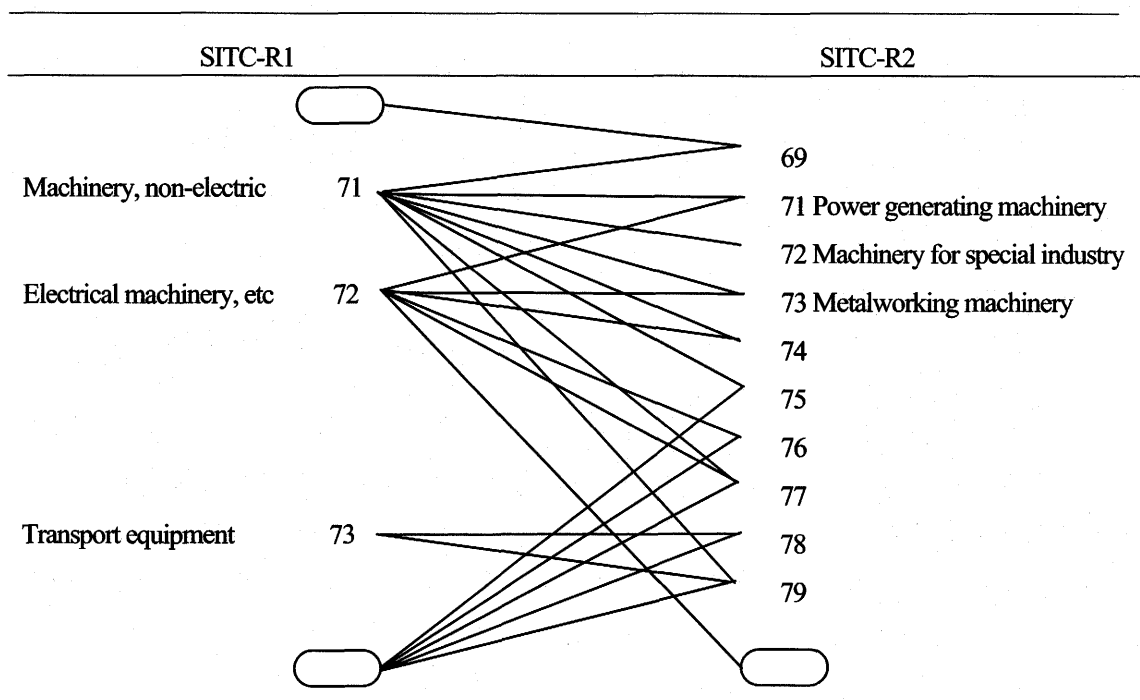
となる^(註7)。ここで、 $m=2, \dots, n-2$ であり、 c_3 は定数とする。

2. 改訂による違いを考慮しない SITC系列の評価

SITCはSITC-R1からSITC-R3まで逐次改訂されているが詳細な商品分類の検討が必要ではないとき、商品分類SITCの1桁レベルあるいは2桁レベルを商品分類の体系による違いを考慮せずに使用することがある。こうした考えは基本項目(最も細かい分類で表わされた商品分類コード)のレベルでは分類による違いが直接的に貿易の取引額に影響されるが、1桁や2桁レベルのような上位桁レベルになると基本項目による違いが互いに相殺し合って表面上は表れなくなるという仮説にもとづいている。この方法は簡便法としてしばしば利用される。しかし、この方法は実態が異なるものがたまたま同一の分類コードで表わされているだけである場合が多いため、本来は薦められない方法である。

第2部の表1にはSITCの1,2,3桁レベルに対して商品分類コードとその名称が改訂版ごとに並べて示されている。この表の1桁レベルでは多くの分類がほぼ似た商品として横並びに見られる。しかし、2桁レベルあるいは3桁レベルになるとその違いが明らかになる。極端な例はSITC-R1とSITC-R2の関係においてSITC-R1には存在しない分類コードがSITC-R2に出現するか、あるいはその逆のような場合である。2桁レベルにおける前者としてはSITC-R1の分類コード74,75,76,77,78,79,87,97、後者はSITC-R2の86があげられる。

図1 分類コード71,72,73の対応関係



(出所) 第2部の表1および表2にもとづいて筆者が作成。

(注) () は明示的に示されていない分類コードの集まりを表す。

同一の分類コードであっても改訂版によってその違いが識別できるものは分類を検討する段階で分類コードの違いが判断される。しかし、問題になるのは類似した名称のときであり、詳細な分類レベルに対する評価あるいは検討を回避するときに簡便法として上記の仮説が設定される。この仮説をOECD貿易統計による日本のデータを利用して検討する。使用した日本のデータは1962年から99年までの期間、輸出入別、相手国は世界合計からなるSITCの1,2,3桁レベルである。

2.1 SITC-R1からSITC-R2への改訂

SITC-R1からSITC-R2への改訂による構造変化は変化点 m と改訂時点の1977が一致しているかどうかで判定する。第2部の表4の「SITC-R1 : SITC-R2」の「1 digit of SITC」には商品分類が1

桁レベルのとき、(13)式の $AIC(m)$ で評価される線形回帰モデルおよび(15)式の $f(m|y)$ で表わされる事後分布で推計された変化点 m が77年と一致した分類コードが示されている。この分類コードは1977年と78年との間に変化点が確認できたことを示しており、すなわち、断層状態が検出されたということの意味する。この表のSeriesは変化点の確認状態を表わし、Xの印は使用した2つのモデルが同一の結果を示していること、*はどちらかのみが一致していることを表わす。Dの#は商品分類の体系としてその分類コードが存在していないかあるいは取引量が少ないかの理由でデータに欠損値があり、1962年から87年までのすべての期間にデータがないことを表わす。欠損値は0と見なされる。すべての期間に渡ってデータがある分類コードにはDに、#が付けられる。Dが#であって同時にEがXである分類コードは多くの場合に前述したような分

類体系上その分類コードのどちらかが存在していないという状態である。

第2部の表4の「Import」よれば、1桁レベルでは輸入における分類コードの9に商品分類の改訂による構造変化が見られることがわかる。すなわち、分類コード9はデータの欠損値の有無を示すDが、であり、同時にEは改訂年が変化点と一致している記号であるXあるいは*が表示されていることから判断される。Eが*であることから2つのモデルのうちどちらかがそれを判断している状態にある要検討の分類コードである。同表において「Export」に1桁レベルの分類コード見られないので、が輸出では改訂年に構造変化がある分類コードは存在しないことになる。

第2章の表4の「SITC-R1 : SITC-R2」 「Import」にある「2 digits level of SITC」より、Dが、であり、同時にXあるいは*であるものが改訂年と変化点が一一致している2桁レベルの分類コードである。輸入における2桁レベルの分類コードは63,71,72,73の4個に分類の改訂年に構造変化が見られる。また、Dが#であり、同時にEがXあるいは*であるものはどちらかの分類コードが存在していない状況を表わす。前述した極端の例である。第2部の表4の2桁レベルおよび3桁レベルをまとめたのが表1である。表1からデータに欠損値がなく改訂年に構造変化を示す4個の分類コード比率は7.1%であることがわかる。

また、これらの4個のコードは第2部の表2によりその対応関係が確かめられる。この対応関係の表をもとにしてSITC-R1の分類コードの71,72,73とSITC-R2のそれとの対応関係を図示したのが図1である。この図からわかるようにSITC-R1の分類コードの71はSITC-R2における69,71,72,73,74,75,77,79の8個のコードと対応しており、SITC-R1の分類コードの72はSITC-R2の71,73,74,76,77,87,88と7個のコードと対応している。また、SITC-R1の73はSITC-R2の78と79の2個と対応している。したがって、この3個の

分類コードそれぞれの取引額の推移を2つのモデルで検証する以前の問題として、対応関係を参照することにより分類体系上の構造変化を事前に確認できる。なお、SITC-R1とSITC-R2の対応関係においてこの3個の分類コードは対応関係のタイプが4bである商品グループ24に属する。同じように63についても確認でき、対応関係が4aである商品グループ10に属する。それ以外ではDに#の印を持つ13個の分類コードにその変化が確認されているが、第2部の表1を参照すればその多くがSITC-R1には存在していないコードであることが確かめられる。

同表の「Export」により、輸出では分類コードの27,52,55,71,72,73,89の7個に構造変化が見られる。分類コードの27,52,89は商品グループ24に属しており、自分自身の分類コード以外に他の分類コードと関係していることが改訂年と変化点の一一致する要因と考えられる。それ以外ではDに#の印を持つ9個の分類コードにその変化が確認されているが、第2部の表1を参照すればSITC-R1あるいはSITC-R2に存在していない分類コードであることが確かめられる。

同じく、「3 digits level of SITC」より3桁レベルでは輸入において2つのモデルの結果が一致して改訂年に構造変化を起こしていると判断される分類コードは15個、輸出は18個であり、分類コードに占める比率としてそれぞれ11.3%と13.5%となる。それ以外に#の印を持つ分類コードにその変化が確認されているが、第2部の表1を参照すればSITC-R1あるいはSITC-R2に存在していない分類コードであることが確かめられる。

2.2 SITC-R2からSITC-R3への改訂

SITC-R2からSITC-R3への改訂による構造変化は変化点*m*と改訂年の1987年が一一致しているかどうかで判定する。第2部の表4の「SITC-R2 :

表1 改訂による違いを考慮しないSITCにおける改訂年の構造変化

系列	種類	SITCの2桁レベル			SITCの3桁レベル					
		AIC(m)	f(m)y	total	AIC(m)	f(m)y	Total			
(1) SITC-R1:SITC-R2										
import	欠損値無し	4 (7.1)	4 (7.1)	56	15 (11.7)	15 (11.7)	128			
	欠損値	9 (69.2)	9 (69.2)	13	152 (96.2)	152 (96.2)	158			
	total	13 (18.8)	13 (18.8)	69	167 (58.4)	167 (58.4)	286			
export	欠損値無し	7 (12.7)	7 (12.7)	55	17 (14.4)	17 (14.4)	118			
	欠損値	9 (64.3)	9 (64.3)	14	152 (90.5)	151 (89.9)	168			
	total	16 (23.2)	16 (23.2)	69	169 (59.1)	168 (58.7)	286			
(2) SITC-R2:SITC-R3										
import	欠損値無し	27 (42.9)	24 (38.1)	63	83 (37.9)	73 (33.3)	219			
	欠損値	2 (50.0)	2 (50.0)	4	51 (94.4)	51 (94.4)	54			
	total	29 (43.3)	26 (38.8)	67	134 (49.1)	124 (45.4)	273			
export	欠損値無し	10 (15.9)	8 (12.7)	63	45 (21.5)	44 (21.1)	209			
	欠損値	2 (50.0)	2 (50.0)	4	49 (76.6)	48 (75.0)	64			
	total	12 (17.9)	10 (14.9)	67	94 (34.4)	92 (33.7)	273			

(出所) 第2部の表4にもとづき筆者作成

(注) () の中の数はパーセントを示す。

表2 商品グループに対する変換の評価

$G_i(j):type D$ Series N	$G_i(j):type D$ Series N	$G_i(j):type D$ Series N
SITC-R1:SITC-R2		
import	15 1 1 . X 1	5 1 1 # 1
	16 1 1 . X 1	7 1 1 # 1
	21 1 1 . X 1	8 1 1 # 1
	22 1 1 . X 1	10 1 1 . X 1
41 1 2 . X 2	32 1 2 . X 1	11 1 1 . X 1
107 1 1 . * 1	33 1 1 # 1	13 1 1 # 1
118 1 1 # 1	34 1 1 . X 1	17 1 4b . X 3
	36 1 1 . X 1	18 1 1 . X 1
export	48 1 1 . X 1	24 1 1 # 1
	49 1 1 . X 1	25 1 1 # 1
12 1 1 . * 1	50 1 1 . X 1	31 1 1 # 1
45 1 1 . X 1	51 1 1 . X 1	33 1 1 # 1
68 1 1 . X 1	53 1 1 . X 1	38 1 1 # 1
103 1 4b . * 130	57 1 1 . X 1	47 1 1 . X 1
118 1 1 # 1	61 1 4b . X 497	54 1 1 . X 1
	65 1 1 . X 1	60 1 1 . X 1
SITC-R2:SITC-R3		
import	77 1 1 # 1	62 1 1 . X 1
	78 1 1 # 1	73 1 1 . X 1
		76 1 1 . X 1
	export	77 1 1 # 1
1 1 1 . X 1		78 1 1 # 1
2 1 1 . X 1	3 1 1 . X 1	

(出所) 筆者作成

(注) 商品グループ $G_i(j):type$ を、 $ij,type$ として表わす。

表3 商品グループにおける改訂年の評価

系列	SITC-R1:SITC-R2			SITC-R2:SITC-R3		
	AIC(m)	$f(m y)$	total	AIC(m)	$f(m y)$	total
(1) 商品グループの個数に対する割合						
import	2 (1.6)	1 (0.8)	122	17 (21.8)	17 (21.8)	78
export	4 (3.3)	2 (1.6)	122	11 (14.1)	11 (14.1)	78
(2) 対応関係の個数に対する割合						
import	3 (1.0)	2 (0.6)	310	17 (2.9)	17 (2.9)	595
export	133 (42.9)	3 (1.0)	310	11 (1.8)	11 (1.8)	595

(出所) 第2部にもとづき筆者作成
 (注) () 中の数はパーセントを示す。

SITC-R3」の「1 digit of SITC」には商品分類が1桁レベルのときの回帰モデルおよび事後分布で推計された変化点 m が87年と一致した分類コードが示されている。この表の「Import」により、1桁レベルの分類コードでは輸入における分類コードの2,4,6,7、「Export」により、輸出では9に改訂による構造変化が見られる。

表1から分類の2桁レベルでは輸入は欠損値がない分類コードに対して63個ある分類コードのうち約40%、欠損値も含めた全体では約30%を改訂年に構造変化を示す分類コードが占める。輸出は67個ある分類コードのうち全体で12個の18%が改訂年に構造変化を示す。同じく、3桁レベルでは273個ある分類コードに対して輸入において全体として改訂年に構造変化を起こしていると判断される分類コードは134個の約50%、輸出は94個の34%を占める。

3. グループ化されたSITC系列の評価

第1章の第2節ではSITCの3桁レベルにおける対応関係グループ化にもとづいて日本の貿易統計について商品グループ化がおこなわれた。本節ではSITC-R1とSITC-R2、SITC-R2とSITC-R3のSITCの3桁レベルにおける対応関係グループ化にもとづく商品グループのそれぞれに対して

改訂年と変化点の一致を検討する。

SITCの3桁レベルにおけるSITC-R1とSITC-R2の対応関係はSITC-R1の183個の分類コードとSITC-R2の239個の分類コードの組合せから構成され、122個の商品グループと310個の対応関係を含んでいる。表2に2つのモデルによる評価の結果が示されている。 $G_i(j):type$ は商品グループ、 D は対象とする期間にすべてのデータが存在していたかどうかを表わし、 $\#$ は欠損値があることを示し、 $.$ はそれがないことを示す。また、 $Series$ は系列を表わし、前述した2つのモデルの評価を示す。この2つのモデルの評価が一致して構造変化を示しているときはX、どちらかるときは*で表示する。 N は商品グループに含まれる対応関係の個数を表わす。

SITC-R1とSITC-R2の対応関係では輸入において商品グループ41と107の2個に変化が見られる。第2部の表3から商品グループ41は対応関係のタイプが2であり、SITC-R1の266とSITC-R2の266および267がグループとして対応していることがわかる。同表から商品グループ107はタイプ1であり、SITC-R1の735とSITC-R2の793が対応している。さらに、表3から改訂年に構造変化が見られたこれらのグループは商品グループに対しては1.6%、対応関係の数に対しては3個の対応関係があるため1.0%を占める。

表4 変換されたSITC各時系列データにおける改訂年の構造変化

商品分類	系列	SITCの2桁レベル			SITCの3桁レベル		
		AIC(m)	f(m/y)	total	AIC(m)	f(m/y)	total
(1) SITC-R1:SITC-R2							
SITC-R1	import: NN	4 (7.0)	4 (7.0)	57	12 (6.8)	11 (6.3)	176
	import: ED	4 (7.0)	4 (7.0)	57	22 (12.5)	20 (11.4)	176
	export: NN	8 (14.3)	7 (12.5)	56	24 (14.5)	22 (13.3)	166
	export: ED	12 (21.4)	12 (21.4)	56	32 (19.3)	30 (18.1)	166
SITC-R2	import: NN	6 (9.4)	5 (7.8)	64	50 (21.5)	48 (20.7)	232
	import: ED	11 (17.2)	10 (15.6)	64	74 (31.9)	70 (30.2)	232
	export: NN	10 (15.9)	10 (15.9)	63	63 (28.5)	63 (28.5)	221
	export: ED	17 (27.0)	17 (27.0)	63	86 (38.9)	85 (38.5)	221
SITC-R3	import: NN	7 (11.1)	7 (11.1)	63	70 (27.5)	65 (25.5)	255
	import: ED	6 (9.5)	6 (9.5)	63	62 (24.3)	57 (22.4)	255
	export: NN	17 (27.4)	17 (27.4)	62	88 (35.8)	87 (35.4)	246
	export: ED	14 (22.6)	13 (21.0)	62	84 (34.1)	83 (33.7)	246
(2) SITC-R2:SITC-R3							
SITC-R1	import: NN	25 (43.1)	21 (36.2)	58	67 (37.6)	62 (34.8)	179
	import: ED	32 (55.2)	28 (48.3)	58	88 (49.4)	84 (47.2)	178
	export: NN	17 (29.2)	15 (25.9)	58	57 (33.7)	56 (33.1)	169
	export: ED	22 (37.9)	23 (39.7)	58	79 (46.7)	81 (47.9)	169
SITC-R2	import: NN	25 (38.5)	23 (35.4)	65	90 (38.5)	84 (35.9)	234
	import: ED	35 (53.8)	30 (46.2)	65	108 (46.2)	103 (44.0)	234
	export: NN	19 (29.2)	18 (27.7)	65	83 (37.1)	83 (37.1)	224
	export: ED	23 (35.4)	25 (38.5)	65	108 (48.2)	109 (48.7)	224
SITC-R3	import: NN	23 (36.5)	21 (33.3)	63	111 (43.5)	103 (40.4)	255
	import: ED	33 (52.4)	29 (46.0)	63	137 (53.7)	133 (52.2)	255
	export: NN	26 (41.3)	22 (34.9)	63	117 (48.1)	114 (46.9)	243
	export: ED	31 (49.2)	30 (47.6)	63	135 (55.6)	131 (53.9)	243

(出所) 第2部の表5にもとづき筆者作成

(注) () の中の数はパーセントを示す。

輸出は商品グループの12,45,68,103に改訂年に構造変化があり、商品グループの12,45,68はタイプ1であるが商品グループ103はタイプ4bである。商品グループの103は130個の対応関係から構成される最大規模のグループである。商品グループおよび対応関係の数の総数は輸入と同じなので、構造変化が見られたグループは商品グループに対しては3.2%、対応関係の数に対しては133個の対応関係があるため42.9%を占める。商品グループ118は輸出入ともにデータに欠損値の存在が確かめられる。

SITC3桁レベルにおけるSITC-R2とSITC-R3

の対応関係はSITC-R2の239個の分類コードとSITC-R3の261個の分類コードの組合せから構成され、78個の商品グループと595個の対応関係を含んでいる。SITC-R2とSITC-R3の対応関係では輸入において17個の商品グループに改訂年に構造変化が見られる。対応関係のタイプはタイプ1が15個、タイプ2とタイプ4bがそれぞれ1個である。タイプ2は商品グループ32であり、4個の対応関係を含み、タイプ4bは最大の規模を持つ商品グループ61であり、497個の対応関係を含んでいる。構造変化が見られたグループは商品グループに対しては21.8%、対応関係の数に対し

ては516個の対応関係があるため86.7%を占める。

輸出は11個の商品グループに改訂年に構造変化が見られる。対応関係のタイプはタイプ1が10個、タイプ4bが1個である。タイプ4bは商品グループ17であり、3個の対応関係を含んでいる。商品グループおよび対応関係の数の総数は輸入と同じなので、変化が見られたグループは商品グループに対しては14.1%、対応関係の数に対しては13個の対応関係があるため2.2%を占める。また、この対応関係にはデータに欠損値を持つ分類コードが存在し、輸入では3個、輸出では11個となっている。

4. ニューラル・ネットワークによるSITC系列の評価

第1章の第3節では日本の貿易統計のSITC3桁レベルにおける分類コードに対して配分ウェイトをもとに1962年から99年までの時系列データとしてSITCの各系列を推計している。本節ではこのSITCの改訂第1版から同第3版の各系列のそれぞれに対して商品分類の改訂年と変化点 m の一致状態を検討する。配分ウェイトの推計を評価するにあたって、比較の系列としてSITC3桁レベルを等配分して作成した系列を使用する。

第2部の表5にSITCの各系列に前述した2つのモデルによる評価の結果が示されている。この表では、 D は対象とする期間のすべてにデータが存在していたかどうかを表し、前節と同じように D において#は欠損値があることを示し、.はそれがないことを示す。 NN はニューラル・ネットワークにより推計された配分ウェイトにより作成された系列、 ED は等分に配分されたウェイトにもとづいて作成された系列を表す。 NN および ED の列に表された E の*は回帰分析あるいは事後分布のどちらかが改訂年と変化点が一致したことを示し、 X はこの2つの方法の結果が

一致して改訂年に構造変化があったことを示す。

この表において最初はSITC-R1の輸入の系列が示される。「1 digit level of SITC」により同1系列における桁レベルの輸入、「2 digits level of SITC」により同2桁レベルの輸入、「3 digits level of SITC」により同3桁レベルの輸入の検討結果が示される。つぎに、SITC-R1の輸出における各系列が同じように示される。続いて、SITC-R2の輸入および輸出、SITC-R3の輸入および輸出の各系列が同じように示される。

SITC-R1の輸入系列におけるSITCの1桁レベルは「1 digit level of SITC」により、SITC-R1とSITC-R2の関係は NN の系列では分類コードの5と9において2つの方法のどちらかが改訂年と変化点が一致する。 ED の系列では分類コードの5が2つの方法が一致して変化の状態を示し、9ではどちらか一方が変化を示す。したがって、この結果から NN と ED ではともに改訂年に構造変化が2個あることになり、配分方法による違いは見られないことになる。SITC-R2とSITC-R3の関係では NN の系列は分類コードの1,6,7、 ED の系列では1,5,6,7が2つの方法が一致して改訂年に構造変化があることを示す。構造変化を示すのは NN が3個に対して ED は4個であるので、わずかなではあるが NN による配分方法のほうがよいといえる。

同じくSITC-R1の輸入系列における2桁レベルはこの表において「2 digits level of SITC」より判断される。SITC-R1とSITC-R2の関係では2つの手法のどちらかが改訂年に変化を示すのが NN では分類コードの29,43,51,95の4個、 ED は09,51,57,63,の4個である。第2部の表5に示されている2桁レベルと3桁レベルをまとめたのが表4である。表4によれば、この NN と ED の4個は7%を占める。SITC-R2とSITC-R3の対応関係ではAIC(m)で評価される線形回帰モデルの結果は NN の系列で25個、 ED の系列で32個であり、そのパーセントはそれぞれ43.1%と55.2%となる。

また、 $f(mly)$ で表される事後分布の結果はNNの系列で21個、EDの系列で28個であり、そのパーセントはそれぞれ36.2%と48.3%である。このことから、NNとEDの違いについてはSITC-R1とSITC-R2の関係においてその違いが見られないものの、SITC-R2とSITC-R3ではNNの方がよいという結果になる。

また、表4の結果より、変化点を推計する2つの方法のうちAIC(m)の線形回帰モデルより $f(mly)$ の事後分布による手法の方が厳しい結果を示している。さらに、表4によれば、SITC-R1の輸入系列の3桁レベルではSITC-R1とSITC-R2の対応関係、SITC-R2とSITC-R3の対応関係ともにNNの方がEDに比べて改訂年における構造変化の数が少なくよい配分方法であるといえる。表4の結果をまとめれば、3桁レベルにおいてNNの方がよい配分方法であると判断されるのはSITC-R1とSITC-R2の対応関係ではSITC-R1の輸出入の両系列、SITC-R2の輸出入の両系列である。しかし、SITC-R3の輸出入の両系列はEDの方がよい配分方法として判断されている。

推計されたSITC-R3の系列は1962年から77年までの分はSITC-R1の原系列を変換してSITC-R2の系列を作成し、さらにこのSITC-R2を再度SITC-R3へ変換することで作成する。最初の変換で作成されたSITC-R2の系列はNNの方がよい配分方法と見なされているので、2度目の変換で不整合な状態が生じたことになる。NNの系列ではSITC-R2からSITC-R3の方向に対する配分ウェイトの推計として前者は1978年から87年までの期間のデータ、後者は88年から93年まで期間のデータを使用している。そのため、原系列の配分構造と推計された配分構造に本質的な違いが生じていたことが想定される。3桁レベルの集計である2桁レベルに対しても同じような結果が得られており、SITC-R1の輸出入、SITC-R2の輸出入の両系列ではNNがよい配分方法であるのに対してSITC-R3の輸出入の両系列はED

の方がよい配分方法として判断されている。

また、SITC-R2とSITC-R3の対応関係ではSITC-R1、SITC-R2、SITC-R3のそれぞれにおける輸出入の両系列、すなわちすべての系列においてがNNの方がよい配分方法であると判断される。3桁レベルの集計である2桁レベルに対しても同じような結果が得られ、すべての系列に対してNNがよい配分方法ということになっている。推計されたSITC-R1の系列は1988年から99年までの分はSITC-R3の原系列を変換してSITC-R2の系列を作成し、さらにこのSITC-R2を再度SITC-R3へ変換することで作成する。SITC-R1とSITC-R2の対応関係のときは違ってSITC-R2とSITC-R3の対応関係ではNNの方がよい配分方法として判断されている。

おわりに

貿易統計を長期時系列データとして利用するには商品分類の改訂や相手国の変更に伴って生ずる不整合の取り扱いが大きな問題となる。本章は商品分類の改訂に伴う不整合の問題に焦点をあて、貿易統計の改訂年前後における不整合の評価方法を示し、それを日本のデータに適用した。

貿易統計は1962年から99年までの時系列データとして、(1) SITCの改訂による違いを考慮せず単にSITCの数字のみを分類コードとみなす系列、(2) グループ化された系列、(3) 配分ウェイトにもとづいて作成された系列、と設定してこの各系列に対して商品分類改訂年の前後に構造変化が生じたかどうかを検討している。

不整合の評価方法にはAIC(m)にもとづく線形回帰モデルおよび $f(mly)$ で表される事後分布を利用して変化点を推計し、商品分類の改訂年と変化点が一致したかどうかを構造変化の判断基準とする。変化点の推計の前にデータの変動

を調整するためにBox-Cox変換を利用する。

SITCの改訂による違いを考慮せずに同一の3桁数字のみを分類コードとして使用したときの系列の結果は第2部の表4に示されており、それをまとめたのが表2である。SITC-R1とSITC-R2の対応関係では輸出入ともに改訂年に約60%の構造変化が見られる。SITC-R2とSITC-R3の対応関係ではそれが輸入で約50%、輸出では約35%を占める。

SITC3桁レベルにおける対応関係グループ化にもとづく商品グループの系列の結果は表3に示されている。商品グループはUNの対応表にもとづいて作成されたものであるため、本来ならば改訂年に構造変化がないはずである。しかし、対応関係がタイプ1であっても構造変化が生じている商品グループがあることに留意する必要がある。

配分ウエイトにもとづいて作成された系列の結果は第2部の表5に示され、それをまとめたものが表5である。ニューラル・ネットワークにより作成された配分ウエイトから作成された系列を*NN*、当配分して作成された系列を*ED*として両者を比較しているが、結果として*NN*の方がよい配分の方法といえる。しかし、*NN*においてもSITCの各系列に対してSITC-R1とSITC-R2の対応関係では約10%から35%、SITC-R2とSITC-R3の対応関係では約35%から50%の割合で改訂年に構造変化が見られる。

繰り返しになるが、変化点が商品分類の改訂年に一致する分類コードあるいは商品グループは必ずしも商品分類の改訂に伴う対応関係の不整合の存在を意味しているわけではない。しかし、不整合であれば変化点が商品分類の改訂年に一致するという状態はおこる可能性がある。こうした観点から本章で紹介した方法は大量に存在する分類コード間の対応関係にある基準に従って改訂年に構造変化が生じていると見なされる分類コードを候補として識別するだけであ

る。したがって、最終的な判断は対応している個別分類コードの例示品目を比較するという検討が必要となる。

(注1) 正確に言えば、多くの国では1994年頃からUN貿易統計あるいはOECD貿易統計は商品分類がHSで分類されるようになってきている。アジア経済研究所が所有しているOECD貿易統計の日本のデータも1993年まではSITC-R3、94年以降はHSの編集である。HSの基本項目(6桁のSub-headingを指す)はSITC-R3へそのまま対応させるかあるいは統合して対応させることにより変換が可能である。SITC-R3とHSの変更の時点を変化点とすると最近年度分のデータが不足するので、本章ではHSをSITC-R3と変換したものをSITC-R3の系列として利用している。

(注2) SITC-R1とSITC-R2の改訂年は1978年であるが回帰分析のさいに年の分割を下記のように設定したため、本章の中では改訂年として1年前の77年として説明している。また、SITC-R2とSITC-R3の改訂年についても1988年を87年としている。「改訂に影響する1962年から1987年までの年別データのみを使用し、この区間のデータを2つに分割する。この分割は、SITC-R1の適用期間を[62,*m*]、SITC-R2の適用期間を[*m*+1,87]と仮定することである。… SITC-R2からSITC-R3への改訂に関しては、改訂に影響する1978年からデータの存在する1993年までのデータのみを使用し、この区間のデータを2つに分割する。この分割は、SITC-R2の適用期間を[78,*m*]、SITC-R3の適用期間を[*m*+1,93]と仮定することである。」

(注3) Box-Cox変換の(2)式において、 λ が0とそれ以外に場合分けされているが、 λ が0のときの極限值は、

$$\lim_{\lambda \rightarrow 0} \frac{y^\lambda - 1}{\lambda} = \log(y)$$

となる。したがって、 $h_j(y)$ は λ に関して連続な関数として定義されていることになる。

(注4) パラメーターが与えられたときの $y_1 \cdots y_m$ の同時分布は、

$$f(y_1 \cdots y_m | \alpha_1, \beta_1, \sigma^2, m) \\ = (2\pi\sigma^2)^{-\frac{m}{2}} e^{-\frac{\sum_{j=1}^m [y_j - (\alpha_1 + \beta_1 x_j)]^2}{2\sigma^2}}$$

である。また、パラメーターが与えられたときの $y_{m+1} \cdots y_n$ の同時分布は、

$$f(y_{m+1} \cdots y_n | \alpha_2, \beta_2, \sigma^2, m) \\ = (2\pi\sigma^2)^{-\frac{n-m}{2}} e^{-\frac{\sum_{j=m+1}^n [y_j - (\alpha_2 + \beta_2 x_j)]^2}{2\sigma^2}}$$

である。そのとき、パラメーターが与えられたときの $y_1 \cdots y_n$ の同時分布は、

$$f(y_1 \cdots y_n | \alpha_1, \beta_1, \alpha_2, \beta_2, \sigma^2, m) \\ = \frac{1}{n-3} (2\pi)^{-\frac{1}{2}} \{\sigma^2\}^{-\frac{n+2}{2}} \cdot \\ \exp \left\{ -\frac{1}{2\sigma^2} \left(\sum_{j=1}^m [y_j - (\alpha_1 + \beta_1 x_j)]^2 \right. \right. \\ \left. \left. + \sum_{j=m+1}^n [y_j - (\alpha_2 + \beta_2 x_j)]^2 \right) \right\}$$

となる。 c_2 は定数とすると、データ $y_1 \cdots y_n$ を観測したときの事後分布の同時密度関数は、

$$f(\alpha_1, \beta_1, \alpha_2, \beta_2, \sigma^2, m | y_1 \cdots y_m) \\ = \frac{f(\alpha_1, \beta_1, \alpha_2, \beta_2, \sigma^2, m)}{f(y_1 \cdots y_m)} \cdot \\ f(y_1 \cdots y_m | \alpha_1, \beta_1, \alpha_2, \beta_2, \sigma^2, m) \\ = c_2 (\sigma^2)^{-\frac{n+2}{2}} \cdot \\ \exp \left\{ -\frac{1}{2\sigma^2} \left(\sum_{j=1}^m [y_j - (\alpha_1 + \beta_1 x_j)]^2 \right. \right. \\ \left. \left. + \sum_{j=1}^m [y_j - (\alpha_2 + \beta_2 x_j)]^2 \right) \right\}$$

と表される。

(注5) この式は θ から ξ への変数変換である。

$$\int e^{-\frac{1}{2\sigma^2}(\hat{\theta}-\theta)'X'X(\hat{\theta}-\theta)} d\theta \\ = \int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2}\xi'(X'X)^{-1}\xi} |(X'X)^{-1}| d\xi$$

$$= |X'X|^{-1} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\xi'[\sigma^2(X'X)]^{-1}\xi} d\xi \\ = |X'X|^{-1} \{(2\pi\sigma^2)^2 |X'X|^{-1}\}^{\frac{1}{2}} \\ = |X'X|^{-\frac{1}{2}} (2\pi\sigma^2)^2$$

この計算には (8) 式の正規分布の性質を利用している。(8) 式にしたがえば、 e の一次変換は4次元のベクトルであり、

$$X'e = \xi \sim N(0, \sigma^2(X'X))$$

となる。これを密度関数で表せば、

$$f(\xi) = (2\pi)^{-\frac{p}{2}} |\sigma^2(X'X)|^{-\frac{1}{2}} \cdot \\ e^{-\frac{1}{2}\xi'[\sigma^2(X'X)]^{-1}\xi}$$

である。密度関数の性質から、

$$\int_{-\infty}^{\infty} e^{-\frac{1}{2}\xi'[\sigma^2(X'X)]^{-1}\xi} d\xi \\ = (2\pi\sigma^2)^{\frac{p}{2}} |X'X|^{-\frac{1}{2}}$$

が得られる。ここで、 $p=4$ である。

(注6) $\Gamma(\alpha)$ はガンマ関数であり、

$$\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x} dx (\alpha > 0)$$

で定義される。

(注7) c_3 を定数とすれば、注5で示された事後分布の同時密度関数から m の周辺密度関数を求めると、

$$f(m | y_1 \cdots y_n) = c_3 \left[m(n-m) \sum_{j=1}^m (y_j - \bar{y}_{j(1,m)})^2 \cdot \right. \\ \left. \sum_{j=m+1}^n (y_j - \bar{y}_{j(m+1,n)})^2 \right]^{\frac{1}{2}} \cdot \\ \left[\sum_{j=1}^m (y_j - \hat{y}_{j(1,m)})^2 + \sum_{j=m+1}^n (y_j - \hat{y}_{j(m+1,n)})^2 \right]^{\frac{n-4}{2}}$$

と表される。これは、

$$\bar{y}_{(1,m)} = \frac{y_1 + \cdots + y_m}{m} \quad \bar{y}_{(m+1,n)} = \frac{y_{m+1} + \cdots + y_n}{n-m}$$

とするとき、

$$\begin{aligned}
|X'X| &= \left[\left(\sum_{j=1}^m y_j \right)^2 - m \sum_{j=1}^m y_j^2 \right] \cdot \\
&\quad \left[\left(\sum_{j=m+1}^n y_j \right)^2 - (n-m) \sum_{j=m+1}^n y_j^2 \right] \\
&= \sum_{j=1}^m (y_j - y_{(1,m)})^2 \cdot \sum_{j=m+1}^n (y_j - y_{(m+1,n)})^2
\end{aligned}$$

となるからである。

(注8) SITC-R1とSITC-R2の対応関係では多くの分類コードの変化点は1973年あるいは74年の第1次オイルショックの前後の激しい経済変動に集中している。第1次オイルショックの影響を取り除くためにモデルに対して次のダミー変数 d_t と d_t^* を追加する。ダミー変数は、

$$d_t = \begin{cases} 1 & t \geq 73 \\ 0 & \text{other} \end{cases} \quad d_t^* = \begin{cases} 1 & t = 73, 74 \\ 0 & \text{other} \end{cases}$$

とする。それを、 $d = (d_1 \cdots d_n)'$ 、 $d^* = (d_1^* \cdots d_n^*)'$ で表す。(7)式の X および θ に対応する行列とベクトルをそれぞれ、

$$\begin{aligned}
X^* &= [X \quad d] & X^{**} &= [X \quad d^*] \\
\theta^* &= \begin{bmatrix} \theta \\ \delta \end{bmatrix} & \theta^{**} &= \begin{bmatrix} \theta \\ \delta^* \end{bmatrix}
\end{aligned}$$

とする。変更した変数にしたがってモデルを分析すると、確かにオイルショックの影響は取り除かれていることはわかるが商品分類の改訂年と変化点が一致する商品グループの変化はそれほど多くはない。したがって、本章ではダミー変数による調整はおこなわないことにする。

【参考文献】

- [1] 赤池弘次「事前分布の選択とその応用」(鈴木雪夫・国友直人編『ベイズ統計学とその応用』東京大学出版会 1989)
- [2] 岩本渉、渡辺千仞、岸野洋久、村上征勝「石油危機と主要製造業の構造変化に関する計量分析」(『統計数理』Vol.38 No.2 統計数理研究所 1990)
- [3] 北川源四郎『FORTRAN 77 時系列解析プログラミング』岩波書店 1993
- [4] 城坂晃正「SITC3桁分類コード変換のための配分ウェイト推計—ニューラル・ネットワークを用いて—」(野田容助編『商品分類の改訂に伴う貿易統計の変換』統計資料シリーズ No.83 アジア経済研究所 2001)
- [5] 鈴木雪夫『統計学』朝倉書店 1993
- [6] 鈴木雪夫「ベイズ統計学の考え方」(鈴木雪夫・国友直人編『ベイズ統計学とその応用』東京大学出版会 1989)
- [7] 野田容助「商品分類の改訂に伴う貿易統計の変換—日本および韓国を例として—」(野田容助編『商品分類の改訂に伴う貿易統計の変換』統計資料シリーズ No.83 アジア経済研究所 2001)
- [8] Diego Salazar, Structural Changes in Time Series Models, *Journal of Econometrics*, Vol.19, No.1, North-Holland, 1982
- [9] Donald Holbert, A Bayesian Analysis of a Switching Linear Model, *Journal of Econometrics*, Vol.19, No.1, North-Holland, 1982