第３章
商品分類の対応関係における配分ウエイトの推計
― SITC-R1 系列の 3 桁レベル分類コード作成に向けて ―

野田容助

はじめに

貿易統計を長期時系列データとして利用するとき商品分類の改訂のおこなわれた年の前後は商品に対する定義およびそのカバレジが必ずしも同じではないため、改訂年を含む連続年では取引金額や数量は整合性のある状態とはいえないことがある。そのため、共通の概念で時系列的に分析をするときには同一商品分類体系へ統一することが必要になる。分類体系統一化は商品分類の改訂年の前後における対応関係にもとづいて配分ウエイトを推計し、このウエイトでそれぞれの分類コードに対応する統計値を再配分、すなわち変換することで可能となる。

本章の目的は OECD 作成による貿易統計データの日本の輸出を利用して SITC 系列の対応関係をもとに配分ウエイトによる変換を通して商品分類体系 SITC-R1 の 3 桁レベル分類コードにもとづく長期時系列データを作成することである。配分ウエイトによる貿易統計の変換は日本と韓国についてはアジア経済研究所統計資料シリーズ No.83 において野田「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ー」と城坂「SITC3 桁分類コード変換のため配分ウエイト推計―ニューラル・ネットワークを用いて―」によるニューラル・ネットワークを推計手法として既に実施されている。ニューラルネットワークの手法は将来的には有効な手法であるとはいえ現在ではまだ高価であり必要に応じて手軽に利用できないという問題を抱えている。したがって、本章では制約条件付き最小 2 乗法により配分ウエイトを推計する方法を採用する。

配分ウエイトは商品グループごとに対応関係の分類コードを基礎として計算されるため商品グループが大きいときには配分ウエイトの数に比して貿易統計データから得られるデータ数が少なくなり推計できなくなることがある。データ数を確保するための方法として本章ではブートストラップ法を採用している。

SITC-R1 系列の 3 桁レベル分類コードに変換するにあたっては直接 3 桁レベル分類コードの対応関係から配分ウエイトを推計するのではなく、4 桁レベル分類コードの対応関係から配分ウエイトを推計して 4 桁レベルの取引金額に変換する。この 4 桁レベル分類コードを集計することで 3 桁レベル分類コードを作成する。

１．対応関係における配分ウエイトの推計

対応関係における配分ウエイトの推計方法は野田の「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ー」を基礎としている。異なる商品分類である分類 A と分類 B の間に相互に関わりを持つ個別分類コードの対応

関係があるとき、分類Aと分類Bの間の閉じた関係を商品グループという。商品グループ内の対応関係の中で、商品分類の改訂に伴って生ずる分類Aから分類Bへの方向に対する変換において分類Aの取引額を配分して分類Bの取引額を推計する方法を示すのが本章の目的である。

商品グループ内の分類 A の m 個ある要素 $A_1 \cdots A_m$ のそれぞれに対する最初の観測値 $x_{11} \cdots x_{m1}$ に分類 B の要素 B_j の最初の観測値 y_{j1} が対応しており、ω_{ij} は分類コード A_i から B_j への方向に対する配分ウエイトとする。観測値から配分ウエイトを推計し、この得られた推計値を利用することにより分類Aから分類Bへの変換が可能となる。分類Aの観測値を配分ウエイトで配分するとき、分類Bの観測値は、

$$y_{j1} = x_{11}\omega_{1j} + \cdots + x_{m1}\omega_{mj} + u_{j1}$$

と表すことができる。ここで、配分ウエイトは $i=1\cdots m$ に対して $\omega_{i1}+\cdots+\omega_{in}=1$ であり、u_{j1} は撹乱項である。分類Bはn個の要素から成っており、$j=1\cdots n$ なので、これを行列表示でまとめると、

$$\begin{pmatrix} y_{11} \\ \vdots \\ y_{n1} \end{pmatrix} = \begin{pmatrix} \omega_{11} & \cdots & \omega_{m1} \\ \vdots & & \vdots \\ \omega_{1n} & \cdots & \omega_{mn} \end{pmatrix} \begin{pmatrix} x_{11} \\ \vdots \\ x_{m1} \end{pmatrix} + \begin{pmatrix} u_{11} \\ \vdots \\ u_{n1} \end{pmatrix}$$

となる。$n\times m$ 行列の配分ウエイト行列を W とおき、さらに、

$$\omega_i' = (\omega_{1i} \cdots \omega_{mi}),\ i=1\cdots n$$

とすれば、配分ウエイト行列は、

$$W = \begin{pmatrix} \omega_{11} & \cdots & \omega_{m1} \\ \vdots & & \vdots \\ \omega_{1n} & \cdots & \omega_{mn} \end{pmatrix} = \begin{pmatrix} \omega_1' \\ \vdots \\ \omega_n' \end{pmatrix}$$

である。また、k個の観測値の存在を考慮して、観測値行列YとXをそれぞれ、$n\times k$ 行列、

$$Y = \begin{pmatrix} y_1' \\ \vdots \\ y_n' \end{pmatrix} = \begin{pmatrix} y_{11} & \cdots & y_{1k} \\ & & \\ y_{n1} & \cdots & y_{nk} \end{pmatrix}$$

と $m\times k$ 行列、

$$X = \begin{pmatrix} x_1' \\ \vdots \\ x_m' \end{pmatrix} = \begin{pmatrix} x_{11} & \cdots & x_{1k} \\ & & \\ x_{m1} & \cdots & x_{mk} \end{pmatrix}$$

とする。撹乱項も同じように $n\times k$ 行列として U とする。すべてに対応関係があるときの配分ウエイト構造は、

(1) $$Y = WX + U$$

と表すことができる。すべての要素が1からなるm次元のベクトルをl_mとすると、(1)式においてウエイトの条件から、

(2) $$W' l_n = l_m$$

が満たされる(注1)。(1) と (2) 式を通常の線形回帰モデルの形式に表現しなおすために (1) 式を転置して、$Y' = X'W' + U'$ とする。この式の各要素をまとめると、$y_i = X'\omega_i + u_i \quad i=1\cdots n$ と表わされ、この式を行列表現で表わせば、

$$\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} X' & & \\ & \ddots & \\ & & X' \end{pmatrix} \begin{pmatrix} \omega_1 \\ \vdots \\ \omega_m \end{pmatrix} + \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix}$$

となる。さらに、ベクトルyを$y' = (y_1' \cdots y_n')$、行列X^*を、

$$X^* = \begin{pmatrix} X' & & \\ & \ddots & \\ & & X' \end{pmatrix}$$

ベクトルωを、

$$\omega = (\omega_1' \cdots \omega_m')'$$

$$= \begin{pmatrix} \begin{pmatrix} \omega_{11} \\ \vdots \\ \omega_{m1} \end{pmatrix} \\ \vdots \\ \begin{pmatrix} \omega_{1n} \\ \vdots \\ \omega_{mn} \end{pmatrix} \end{pmatrix}$$

ベクトル u を $u' = (u_1' \cdots u_n')$ とする。これをまとめると、

(3) $$y = X^*\omega + u$$

となる。m 次元の単位行列を I_m として、I_m を n 個横に並べた行列を、$C = (I_m \cdots I_m)$ とする。

ベクトルωはウエイトを表わしているので、

(4) $$C\omega = \omega_1 + \cdots + \omega_n = l_m$$

となる。

1.1 商品グループの配分ウエイト構造

商品分類の対応関係では商品グループ内に対応関係がないウエイトが存在するのが一般的であり、ω_i の要素の中に 0 となるものが含まれる。(3) 式および (4) 式はω_i の要素の中に 0 が含まれていないという前提で作られているが、この部分に 0 も含むように設定を変更する必要がある。配分ウエイトω_i においてその要素が 0 であるものを取り除き、0 でない要素のみからなるベクトルをつくる。ω_i の 0 である要素を ω_{ij} とするとき、m 次元の単位行列から j 行を取り除いた行列を D_i とする。取り除く j は複数個あってもかまわない。

要素として 0 を含まないベクトルを調整済みベクトルとする。調整済みベクトルは 1 次変換により、$\omega_i^D = D_i\omega_i$, $i = 1 \cdots m$ として得ることができる。ω_i^D に対して $D_i = I_m$ であるので、

$$(\omega^D)' = ((\omega_1^D)' \cdots (\omega_m^D)')$$

と、

$$D = \begin{pmatrix} D_1 & & \\ & \ddots & \\ & & D_m \end{pmatrix}$$

とすれば、$\omega^D = D\omega$ となる。

ベクトルωに対応する観測値行列 X^*および行列Cに対しても同じような操作が必要となる。すなわち、ωから取り除く要素に対応する X^* のすべての列を取り除き調整済みの行列として、$X^{*D} = X^* D'$ とする。また、同じようにして C についても調整済み行列は、$C^D = CD'$ とする。なお、C^D に関してはすべてが 0 である要素からなる行は削除する。

対応関係の調整済みのω^D、X^{*D} と C^D に対して、(3) 式に相当するのが、

(3') $$y = X^{*D}\omega^D + u^D$$

であり、(4) 式に相当するのが、

(4') $$C^D\omega^D = l_m$$

である。

1.2 制約条件付き最小 2 乗法による推計

配分ウエイトの推計にはラグランジェ乗数法を利用して、(4') 式を満足するという制約条件のもとで (3') 式の残差平方和$(u^D)'u^D$ を最小にするω^D の値$\tilde{\omega}^D$ を求めるという方法を採用する。ベクトルをλ とおいて、スカラーs を次のようにする。

(5) $$s = (u^D)'u^D + \lambda'(C^D\omega^D - l_m)$$

この式に (3') 式を代入した後、ω^D に関して偏微分して 0 とおく。

$$\partial s / \partial\omega^D = -2(X^{*D})'(y - X^{*D}\omega^D) + (C^D)'\lambda$$

この式より、$(X^{*D})'X^{*D}$ 行列が正則行列であれば$\tilde{\omega}^D$ が得られ、

(6) $$\tilde{\omega}^D = \hat{\omega}^D - [(X^{*D})'X^{*D}]^{-1}(C^D)'\lambda$$

となる。ここで、$\hat{\omega}^D$ は (5) 式において (4') 式の制約条件がないときに得られるω^D の最小 2 乗推定量であり、

(7) $$\hat{\omega}^D = [(X^{*D})'X^{*D}]^{-1}(X^{*D})'y$$

である。

さらに、$\partial s / \partial\lambda = 0$ を満足する$\tilde{\omega}^D$ は、

(8) $$C^D\tilde{\omega}^D - l_m = 0$$

も同時に満足する必要がある。これは (4') 式の制約条件そのものである。(6) 式の両辺に右から C^Dをかけ、(8) 式を利用すれば、

$$C^D\tilde{\omega}^D = C^D\hat{\omega}^D - C^D[(X^{*D})'X^{*D}]^{-1}(C^D)'\lambda$$
$$= l_m$$

となる。この式からλ を求めて、

$$\lambda = \{C^D[(X^{*D})'X^{*D}]^{-1}(C^D)'\}^{-1} \bullet$$
$$(C^D\hat{\omega}^D - l_m)$$

を得る。さらに、

(9) $$M=[(X^{*D})'X^{*D}]^{-1}(C^D)'\bullet\{C^D[(X^{*D})'X^{*D}]^{-1}(C^D)'\}^{-1}$$

とするとき、λを (6) 式に代入して整理し、(9) 式を利用すれば、制約条件付きの最小2乗推定量、

(10) $$\begin{aligned}\tilde{\omega}^D&=\hat{\omega}^D-M(C^D\hat{\omega}^D-l_m)\\&=(I-MC^D)\hat{\omega}^D+Ml_m\end{aligned}$$

が得られる。

2．タイプごとの配分ウエイト

前述したように、分類Aと分類Bの対応関係において、分類Aから分類Bの方向に対する変換を前提として、取引金額のXとYは実額ではなく比率で表されているとする。

対応関係にある分類Aの個数mと分類Bの個数nに対して、$m=n=1$は対応関係のタイプ1である。配分ウエイトはω_{11}しか存在せず、(4) 式において C=1 なので$\omega_{11}=1$である。対応関係のタイプ1のウエイトは1である。

分類Aの個数が m=1 であり、分類Bの個数nが1ではないときは対応関係のタイプ2である。配分ウエイトは$\omega=(\omega_{11}\ \omega_{12}\cdots\omega_{1n})'$のみが存在しており、$C=(1 1\cdots 1)$である(注2)。$m$=1 であるため、$x_1=(1\cdots1)'=l_k$であり、

$$(X^{*}{}'X^{*})^{-1}=\begin{pmatrix}l_k'l_k&&\\&\ddots&\\&&l_k'l_k\end{pmatrix}^{-1}=I/k$$

となり、(9) 式より、$M=(1\cdots1)'/n$となる。また、(7) 式より、$\hat{\omega}^D=(l_k'y_1\cdots l_k'y_n)'/k$となるので、

$$\tilde{\omega}^D=\left(\begin{pmatrix}1\\\vdots\\1\end{pmatrix}+\begin{pmatrix}n-1&\cdots&-1\\\vdots&\ddots&\\-1&\cdots&n-1\end{pmatrix}\hat{\omega}^D\right)/n$$

である。$\tilde{\omega}^D$の最初の要素ω_{11}は、

$$\omega_{11}=\{1+l_k'[(n-1)y_1-(y_2+\cdots+y_n)]/k\}/n$$

であり、$x_1=y_1+y_2+\cdots+y_n=l_k$となるので、

$$\begin{aligned}\omega_{11}&=\{1+l_k'(ny_1-l_k)/k\}/n\\&=l_k'y_1/k=\bar{y}_1\end{aligned}$$

となる。$\bar{y}_1$はy_1の平均である。同じようにして、$j=1\cdots n$に対して、$\omega_{1j}=l_k'y_j/k=\bar{y}_j$となる。$\bar{y}_j$は$y_j$の平均である。対応関係のタイプ2のウエイトは平均により割り振られる。

分類Bの個数が n=1 であり、分類Aの個数mが1ではないときは対応関係のタイプ3である。配分ウエイトは$\omega_1=(\omega_{11}\ \omega_{21}\cdots\omega_{m1})'$のみが存在する。$C=I_m$となるので、$C\omega_1=\omega_1=\ell_m$が得られる。したがって、対応関係のタイプ3のウエイトはすべて1となる。

mとnがともに1より大きく、配分ウエイトの個数が m+n-1 のときは対応関係のタイプ 4a である。配分ウエイト行列として、

$$W=\begin{pmatrix}\omega_{11}&\omega_{21}&\cdots&\omega_{m1}\\\omega_{12}&0&\cdots&0\\\vdots&\vdots&&\vdots\\\omega_{1n}&0&\cdots&0\end{pmatrix}$$

であると仮定すると、配分ウエイトは、

$$\omega^D=((\omega_{11}\cdots\omega_{m1})'\ \omega_{12}\cdots\omega_{1n})'$$

となる。これに対応するDは、i番目の要素が1でそれ以外の要素がすべて0であるベクトルを$e_i=(0\cdots010\cdots0)$とすれば、

$$D=\begin{pmatrix}I_m&&&\\&e_1'&&\\&&\ddots&\\&&&e_1'\end{pmatrix}$$

と表される。さらに、調整済みのCを計算して、

$$\begin{aligned}C^D&=CD'=(I_m\cdots I_m)D'\\&=(I_m\ e_1\cdots e_1)\\&=\begin{pmatrix}1&&&1&&1\\&1&&0&&0\\&&\ddots&\vdots&\cdots&\vdots\\&&1&0&&0\end{pmatrix}\end{aligned}$$

なので、(4') 式を利用すれば、

(11) $$C^D\omega^D=\begin{pmatrix}\omega_{11}+\cdots+\omega_{1n}\\ \omega_{21}\\ \vdots\\ \omega_{m1}\end{pmatrix}=l_{m+n-1}$$

が得られる。調整済みの X^* は、

$$X^{*D}=X^*D'=\begin{pmatrix}X' & & & \\ & x_1 & & \\ & & \ddots & \\ & & & x_1\end{pmatrix}$$

であり、X^{*D} と $(\omega_{21}\cdots\omega_{m1})=(1\cdots1)$ を（3'）式へ代入して、

(12) $$\begin{pmatrix}y_1\\ y_2\\ \vdots\\ y_n\end{pmatrix}=\begin{pmatrix}X' & & & \\ & x_1 & & \\ & & \ddots & \\ & & & x_1\end{pmatrix}\begin{pmatrix}\omega_1\\ \omega_{21}\\ \vdots\\ \omega_{m1}\end{pmatrix}+\begin{pmatrix}u_1\\ u_2\\ \vdots\\ u_n\end{pmatrix}$$

となる。(12) 式の1行目だけを取り出せば、

$$y_1=X'\omega_1+u_1=(x_1\cdots x_m)(\omega_{11}\ 1\cdots1)'+u_1$$
$$=x_1\omega_{11}+X'(0\ 1\ \cdots1)'+u_1$$

である。$v=y_1-X'(0\,1\cdots1)'=x_1\omega_{11}+u_1$ とおいて、(12) 式の y_1 を v に置き換えれば

$$\begin{pmatrix}v\\ y_2\\ \vdots\\ y_n\end{pmatrix}=\begin{pmatrix}x_1 & & & \\ & x_1 & & \\ & & \ddots & \\ & & & x_1\end{pmatrix}\begin{pmatrix}\omega_{11}\\ \omega_{12}\\ \vdots\\ \omega_{1n}\end{pmatrix}+\begin{pmatrix}u_1\\ u_2\\ \vdots\\ u_n\end{pmatrix}$$

となる。この式は $m=1$ であり、n が1ではないときの対応関係であり、対応関係のタイプ2となる。

対応関係のタイプがタイプ1、タイプ2、タイプ3、タイプ4a のいずれでもないのがタイプ4b である。このときの配分ウエイトは（10）式で得られる。

3．疑似ブートストラップ法による標本抽出

推定量 $\hat{\theta}$ の標本分布および標本誤差は母集団から反復抽出された標本に対する推定値の変動を表わす。実際のデータ解析では母集団から抽出された標本がただ1組あるだけでそれにもとづいて推定値を求めると同時に推定誤差を評価しなければならない。この1組の標本を用いて実験的に母集団から標本を反復抽出するには母集団の確率分布が何らかの方法で特定できれば可能となる。

未知の分布関数 $F(x)$ をもつ母集団から大きさ n の無作為標本にもとづく推定量 $\hat{\theta}$ の標本分布は $F(x)$ が知られていないため直接導出できない。そこで、$F(x)$ を経験分布関数 $\hat{F}(x)$ で推定して、経験分布関数を母集団分布としてもつ母集団からの無作為標本にもとづく推定量 $\hat{\theta}*$ の分布で推定量 $\hat{\theta}$ の分布の近似をおこなう。

［1］観測された n の標本を $x_1\cdots x_n$ とし、標本のそれぞれに $1/n$ の確率を付与することにより経験分布関数 $\hat{F}(x)$ をつくる。

［2］経験分布関数 $\hat{F}(x)$ を持つ母集団から抽出された大きさ n の無作為標本を $X_1^*\cdots X_n^*$ とする。これをブートストラップ標本という。

［3］ブートストラップ標本をもとに推定量 $\hat{\theta}$ の標本分布を、

$$\hat{\theta}^*=\hat{\theta}^*(X_1^*,\cdots,X_n^*)$$

の分布で近似する。$\hat{\theta}^*$ の分布をブートストラップ分布という。

多くの推定量に対しては標準誤差を含めてその分布を理論的に導くことはきわめて困難である。そこで、コンピュータを使用して観測された標本に基づく反復抽出という［1］から［3］のプロセスを通して推定量 $\hat{\theta}*$ の分布および標準誤差を求める。

確率分布モデルの分布型だけが仮定できるような場合には母集団にその確率分布モデルを想定する。母集団から得られた標本により分布を規定するパラメターを推定値で置き換えることでモデルを特定し、その確率分布に従う乱数を

反復発生させて推定量の確率的変動をとらえようとするのがパラメトリック・ブートストラップ法である。分布型に関する情報もない場合には、母集団から得られた標本により分布関数を推計する必要がある。この推計された分布関数を経験分布関数といい、経験分布関数をもつ母集団から標本の反復抽出により推定量を評価するのがノンパラメトリック・ブートストラップ法である。

本章では確率分布の分布型は正規分布であることが知られているが、分布を規定するパラメターの値は知られていないとして、標本推定誤差をパラメトリック・ブートストラップ法により評価する方法を用いる。確率分布モデルが特定されると母集団からの標本抽出過程はコンピュータを用いて特定した分布に従う乱数を反復発生させることによって実験的におこなうことができる。

［1］観測されたn個のデータを用いて標本平均値$\bar{x}$、標本分散値s^2を計算する。

［2］n個の標準正規乱数$z_1 \cdots z_n$を発生させ、［1］で求めた標本平均値と標本分散値から、

$$x_i = \bar{x} + sz_i \quad i = 1 \cdots n$$

と変換する。

［3］［2］で発生させた大きさnの標本にもとづいて標本平均を計算し、これを$\bar{x}_1$とする。同じようにして［2］により新たに大きさがnの標本を発生させてその平均値を$\bar{x}_2$とする。このプロセスをM回繰り返して$\bar{x}_1 \cdots \bar{x}_M$を得る。

［4］［3］で求めたM個の標本平均値に基づいて推定量$\bar{X}$の標本誤差の2乗は、

$$sd(\bar{X})^2 = \sum_{i=1}^{M} (\bar{x}_i - \bar{x}_{(\bullet)})^2 /(M-1)$$

で推定される。ここで、

$$\bar{x}_{(\bullet)} = \sum_{i=1}^{M} \bar{x}_i / M$$

としている。推定値$\bar{X}$の分布関数の値は、$P(\bar{X} \leq x) = \{M$個の$\bar{x}_i \; i = 1 \cdots M$ のうち、xより小さいか等しい$\bar{x}_i$の個数$\}/M$で近似される。本章では分布が正規分布と仮定されているので分布関数の近似を利用せず、直接正規乱数を使用する「疑似」ブートストラップ法を採用する。

４．商品分類 SITC-R1 系列の作成

本章で使用される貿易統計データはアジア経済研究所が所有している世界貿易統計データベースシステム：AID-XT（Ajiken Indicators on Developing Economies: Extended for Trade Statistics）を基礎データとしており、日本の輸出はOECD 作成の貿易統計であり、2003 年 3 月現在では1962年から2000年までをカバーしている。本章ではこの中から 1962 年から 99 年までを対象とする（注3）。

アジア経済研究所では国際機関から入手した貿易統計データについて報告国、輸出入区分ごとに商品分類および相手国のサムチェックによる整合性の検討をおこない、商品分類に関しては整合性が保証されていない国はできるだけサムチェックという意味において整合性のあるようにデータの補正をしている（注4）。また、商品分類コードの中で下位のレベルの分類コードを持たないものを詳細分類コードといい、詳細分類コードの取引額を合計すると商品総額に一致する分類コードの集まりを整合性のある詳細分類コードという。アジア経済研究所ではこの整合性のある詳細分類コードによる商品分類をもとにした貿易統計データを AID-XT の基礎データとしている。日本の 1962 年から 1999 年までの輸出データにおける SITC-R1 系列の作成にあたっては AID-XT の基礎データをもとに以下の変換が必要である。

［1］日本の輸出データにおいて、1962年から77年までの範囲はSITC-R1の原系列が存在する

のでこれを使用する。

［2］1978年から87年までは原系列はSITC-R2であるのでSITC-R2からSITC-R1への方向を持つ配分ウエイトを推計してSITC-R1の系列を作成する。

［3］1988年から93年までは原系列はSITC-R3なのでSITC-R3からSITC-R2を推計し、続いてこの得られたSITC-R2の系列をSITC-R1へと再度変換する。

［4］1994年以降の原系列はHSであるため、SITC-R3へ変換した後上記の方法を適用する。

本章で対象とする日本の貿易統計データにおいて1994年から99年までは商品分類がHS-O（1988年度版）であるのでHS-OとSITC-R3の対応表にしたがってSITC-R3へと変換する。HS-OからSITC-R3の対応関係はHS-Oの分類コードに対して対応関係のタイプ1が2,218個、対応関係のタイプ3の個数が2,799個、これは商品グループにすれば893個に当たるものが存在している他にタイプ4aで示される配分構造を持つ商品グループが1つ存在する。この商品グループの変換はデータ処理の複雑さを取り除くため、均等配分による方法を適用する(注5)。この変換された系列を集めて1962年から99年までの期間におけるSITC-R1の系列の時系列データが得られる。

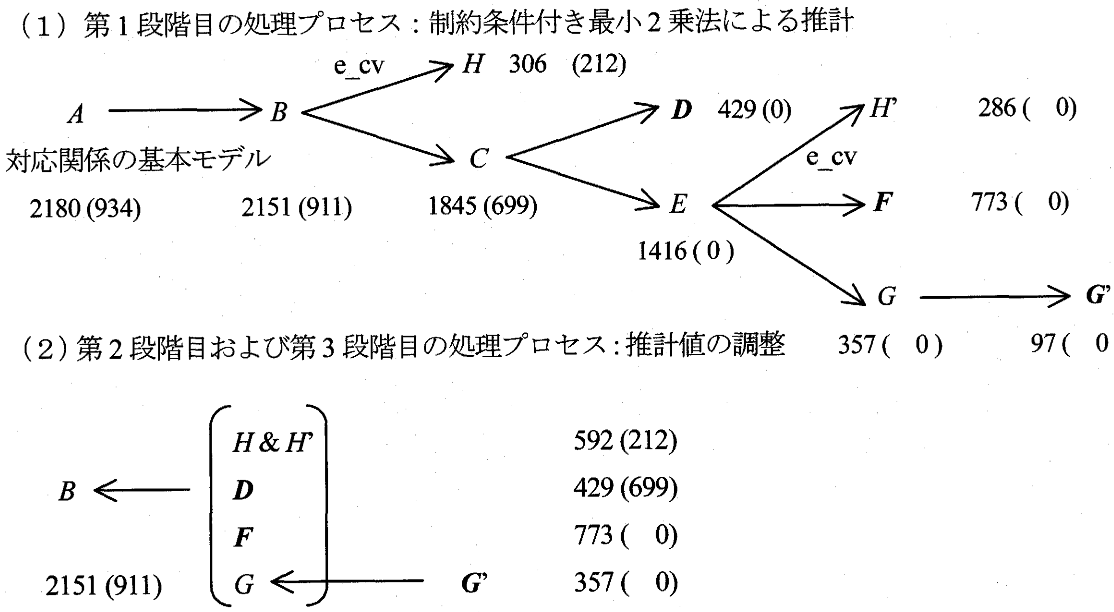
4.1 商品分類の対応関係と変換

商品分類SITC-R1とSITC-R2の基本項目による対応関係コード表はUN統計局発行の*Standard International Trade Classification, Revision 2*、商品分類SITC-R2とSITC-R3の基本項目における対応関係コード表はUN統計局発行の*Standard International Trade Classification Revision 3*から得られた対応関係を対応関係の基本モデルとする。本章では4桁レベルの分類コードをもとにして変換するため、変換にはこの基本モデルから4桁分類コードの対応関係を作成したものを利用する。HS-OとSITC-R3の対応関係コード表は同じくUN統計局発行の*Commodity Indexes for the Standard International Trade Classification, Revision 3*から得られた対応表を基本モデルとして利用する。

日本の輸出における配分ウエイトの推計プロセスが図1に示されている。この図の*A*が商品分類の分類体系から得られた4桁レベルの分類コードにおける対応関係の基本モデルである。SITC-R2からSITC-R1への変換のための対応関係の個数が（　）の中の数字で示され、934個あることがわかる(注6)。SITC-R3からSITC-R2への変換のための対応関係の個数は（　）なしの数字で示され、2,180個である。この対応関係の基本モデルを出発点として配分ウエイトを推計していくが、対応関係のタイプ1とタイプ3に属する分類コードの配分ウエイトは無条件に1であるので推計はおこなわず、タイプが2,4a,4bのみを対象として配分ウエイトを推計する。

配分ウエイトの推計は3段階の処理プロセスから構成される。第1段階目のプロセスは計算可能な商品グループに対して制約条件付き最小2乗法により推定量を計算することである。計算プログラムの内部容量の関係から推計するパラメターである配分ウエイトの数が多すぎて計算不可能な大きな商品グループは予め取り除くておく必要がある。取り除いた商品グループは切断により比較的小さなサブグループを作成することができる(注7)。サブグループが計算可能であれば制約条件付き最小2乗法により推定量を計算する。サブグループにおいても大きすぎるものが残ることがあるが、そのときは4桁レベルの分類コードの対応関係ではなく3桁レベルの分類コードの対応関係に直して計算する。日本の輸出データではこの段階ですべての商品グル

図1　日本の輸出データにおける配分ウエイトの推計プロセス



（出所）著者作成

（注）e_cv は切断によって得られる切断の要素を示す。（　）なしの数字は SITC-R3 から SITC-R2 の対応関係の個数であり、（　）の中の数字は SITC-R2 から SITC-R1 への対応関係の個数を表す。(2) において、D の箇所に(699)となっているが、SITC-R2 と SITC-R1 の対応関係では C に相当する個数である。

ープが推計される。

第1段階目の配分ウエイトの推計プロセスは次のようにまとめられる。

［1］各国の貿易取引状況を考慮せずに分類体系のみから得られる4桁レベルの分類コードの対応関係をAとする。

［2］対応関係のAは一般的な対応関係であるので、このAから日本の輸出データで使用されているすべての分類コードのみを対象とした対応関係を作成する。この対応関係が図1における対応関係のBである。本章の目的は対応関係Bの配分ウエイトを推計することである。この対応関係Bが商品分類の変換のときに利用される。

［3］日本の輸出データの中で取引金額が500US$以上であり、同時に5年以上取引実績のある分類コードのみを対応関係Bから取り出し対応関係Cを作成する。この対応関係の作成は上記の条件を満たさない対応関係を切断の要素Hとするときの切断によるサブグループ化に相当する。対応関係Cは配分ウエイトの推計のための商品グループのみから構成される計算可能な対応関係である。

［4］日本の輸出データと対応関係Cから配分ウエイトを推計する。データ処理の煩雑さを避けるため対応関係のタイプ2、4a、4bのみを推計の対象とする。

［5］対応関係Cにおいてすべての商品グループが推計できれば第1段階目のデータ処理は終了である。しかし、多くの場合、推計プログラムの容量の関係から商品グループの大きさがある程度を超えたときには計算できない。そのときは対応関係Cの中から商品グループの大きな

対応関係*E*とそうではない対応関係*D*とに分割する必要がある。

［6］対応関係*D*に含まれる商品グループは計算可能であるのでの配分ウエイトを推計する。

［7］計算不可能な対応関係*E*は対応関係を切断して計算可能な大きさのサブグループを作成する。実は本章の配分ウエイトの推計作業においてこの切断の要素*H'*を確定することが一番厄介な作業である。

［8］切断によりサブグループが作成されるがこの中にも［5］と同じ問題が含まれる。したがって、対応関係の*E*は計算可能な対応関係*F*とそうではない対応関係*G*と切断の要素*H'*に分割する。

［9］対応関係*F*に含まれる商品グループの配分ウエイトを推計する。

［10］対応関係*G*はこれ以上の切断はおこなわずに、3桁レベルの分類コードの対応関係*G'*に置き換えて少しでも対応関係の大きさを計算可能な状態に直して推計する。これが推計できれば配分ウエイトのデータ処理の第1段階目は終了である。

［11］分類コードの3桁レベルで得られた配分ウエイトの推計値を4桁レベルへ再配分する。この方法としては1個の3桁レベルの分類コードを複数個の4桁レベルの分類コードに配分することから対応関係のタイプ2の方法で処理できる。しかし、本章ではデータ処理を簡単にするため均等配分の方法を適用している。これは、図1において (2) の*G'*から*G*に向かう矢印に対応する箇所である。

［12］日本の輸出においては幸いにも［9］のプロセスで終了したが、そうではない場合の処理方法については今後の検討課題である。

第1段階目で対応関係の*D,F,G*に属する4桁レベルの分類コードの配分ウエイトが推計される。推計された対応関係は図1において太字にて表されている。SITC-R2とSITC-R1の関係については第1段階目は対応関係の*C*、SITC-R3とSITC-R2の関係は対応関係の*G'*で第1段階目が終了する。

4.2 配分ウエイトにおける負の調整

商品分類の改訂年前後の貿易構造によっては商品グループ内の配分ウエイトは必ずしも正の値を取るとは限らない。例えば、商品グループによっては負の配分ウエイトが推計されることがある。この負の配分ウエイトに対する変換元の取引額が大きいと負の取引額が得られる。したがって、配分ウエイト負の値が含まれているときには、すべての配分ウエイトが正の値となるように再配分をする必要がある。この再配分による調整を配分ウエイトの「負の調整」といい､配分ウエイトの推計のための第2段階目のプロセスである。

負の調整は配分ウエイト$\omega_1 \cdots \omega_n$に負の値が含まれているとき、最小の値ω_1が0.0001となるようにすべての配分ウエイトから$\omega_1 - 0.0001$を差し引き､その合計で配分ウエイトを除する。すなわち、$i = 1 \cdots n$に対して、

$\omega_i' = \{\omega_i - (\omega_1 - 0.0001)\} / \{\omega_{\bullet} - n(\omega_1 - 0.0001)\}$

となるのが負の調整済みの配分ウエイト$\omega_1' \cdots \omega_n'$である。$\omega_{\bullet}$は$n$個の合計を表す。したがって、$\omega_1'$は必ずしも置き換えた値の0.0001には等しくない。

4.3 配分ウエイトの最終調整

配分ウエイト推計の第3段階目は切断の要素として第1段階目の処理から取り除いた対応関係の推計である。野田・山本の「体系の異なる分類の対応関係と変換－グループ化と切断による商品分類の試み－」で述べているように、切

表1　日本の輸出データにおける対応関係のタイプが2,4a,4bにおける配分ウエイトの推計値

G	*XA*	*XB*	*A_f*	*B_f*	*ω*	*e*
SITC-R2→SITC-R1						
017	0224	0221	2	1	0.50000	0
017	0224	0222	2	1	0.50000	0
037	0484	0484	2	1	0.93885	1
037	0484	0488	2	2	0.06115	1
037	0488	0488	1	2	1.00000	1
042	0575	0515	2	1	0.50000	0
042	0575	0520	2	2	0.50000	0
042	0579	0519	2	1	0.99990	1
042	0579	0520	2	2	0.00010	0
089	2119	2118	2	1	0.00010	0
089	2119	2119	2	1	0.99990	1
097	2332	2313	2	1	0.00010	0
097	2332	2314	2	1	0.99990	1
098	2450	2411	2	2	0.50000	0
098	2450	2412	2	1	0.50000	0
098	2460	2411	3	2	0.33333	0
098	2460	2421	3	1	0.33333	0
098	2460	6318	3	4	0.33333	0
098	6349	6318	1	4	1.00000	1
098	6352	6318	2	4	0.50000	0
098	6352	6322	2	1	0.50000	0
098	6359	6318	2	4	0.00009	1
098	6359	6328	2	1	0.99990	1
101	2479	2424	2	1	0.50000	0
101	2479	2429	2	1	0.50000	0
108	2519	2515	2	1	0.50000	0
108	2519	2519	2	1	0.50000	0
112	2614	2611	2	1	0.00010	0
112	2614	2612	2	1	0.99990	1
117	2686	2626	2	1	0.00010	0
117	2686	2629	2	1	0.99990	1
119	6512	2628	2	1	0.22784	1
119	6512	6512	2	1	0.77216	1
124	2640	2640	2	1	0.50000	0
124	2640	2658	2	2	0.50000	0
124	2659	2658	1	2	1.00000	0
125	2651	2651	2	1	0.50000	0
125	2651	2653	2	1	0.50000	0
142	2772	2752	1	2	1.00000	1
142	2789	2752	3	2	0.34185	1
:						

(出所) 筆者作成

断の要素は「… 機械的に切断する方法を取らずにグループ間の分類の内容を個別に検討して比較的関係がなさそうだと判断される対応関係を切断する方法を採用する。」という特性を持っている。すなわち、配分ウエイトの値が小であることを仮定しているわけである。したがって、切断の要素の配分ウエイトを同一商品分類の中の第2段階目までで得られた最小の値に0.0001乗じた値を初期値として設定する。最終的に配分ウエイトはそれらの合計が1となるように調整された値である。

推計された4桁レベル分類コードの配分ウエイトは第3部の表1に掲載されている。この表の一部が本章の表1である。表1では紙面の関係から対応関係のタイプ2,4a,4bに属する分類コードのみしか示していないが、第2節で示したようにそれ以外の配分ウエイトはすべて1である。表1において*G*は商品グループの一連番号を示す。*XA*は変換元の4桁分類コード、*XB*は変換先の4桁レベルの分類コード、*A_f*は分類コード*XA*の頻度、*B_f*は分類コード*XB*の頻度を表す。*ω*は分類コード*XA*から*XB*への配分ウエイトである。SITC-R2からSITC-R1の変換をSITC-R2→SITC-R1で表し、*XA*がSITC-R2、*XB*がSITC-R1にそれぞれ対応する。同じようにSITC-R3からSITC-R2の変換はSITC-R3→SITC-R2となり、*XA*がSITC-R3、*XB*がSITC-R2となる。

表1の*e*は配分ウエイトの計算可能状態を表し、1は計算可能であり制約条件付きの最小2乗法により得られた推定値、0は計算不可能のため均等配分した推計値であることを示す。例えば、商品グループ：017では対応関係のタイプが2であり、SITC-R2の0224がSITC-R1の0221と0222の2つに配分されるが、変換先の*e*は共に0でありデータの欠如により計算不可能なため均等配分により0.5としてウエイトが得られている。商品グループ：037はSITC-R2の0484と0488がSITC-R1の0484と0488に対応するタイプ4aの関係にある。この商品グループに属する分類コードのすべての*e*は1であるので計算により求められた配分ウエイトである。

同一商品グループに属している対応関係で*e*が0と1の両方を含んでいるときは、最初に*e*が1の対応関係のみを使用して配分ウエイトを計算

表２　日本の輸出データにおいて変換により断層が生じていると想定される商品分類コード

C	*DW*	*UW*
1978		
012	2.000*	2.000*
013	4.919*	0.489
022	7.636*	7.591*
023	19.000*	19.000*
025	1.583	4.155*
042	4.244*	4.244*
052	0.103	1.793
099	0.213	1.300
241	2.567*	4.333*
244	0.244	0.244
261	2.902*	2.902*
262	2.504*	0.541
264	2.061*	2.061*
265	2.272*	2.263*
276	1.556	2.562*
285	4.176*	4.176*
291	2.119*	0.964
341	5.177*	5.177*
521	1.458	14.277*
551	1.352	2.009*
561	1.236	0.385
656	2.270*	0.892
733	3.200*	2.363*
734	2.436*	2.451*
842	2.405*	2.405*
1988		
011	1.174	5.815*
012	0.528	153.014*
013	0.805	35.521*
022	0.481	0.754
023	6.667*	6.667*
043	2.500*	2.500*
044	.	1.167
047	0.202	0.202
048	1.845	2.091*
052	0.045	0.606
053	0.857	3.215*
054	0.359	0.309
073	8.419*	4.761*
091	0.354	0.335
111	0.405	0.394
121	0.392	0.392
122	2.256*	2.256*
221	11.790*	11.790*
243	0.267	1.014
262	1.117	2.382*
264	5.000*	5.000*
331	0.533	0.385
411	2.882*	1.438
421	0.336	7.560*
422	0.458	1.083
431	0.921	2.384*
515	3.147*	7.349*
613	0.433	0.433
621	2.110*	3.308*
632	3.742*	2.859*
652	0.352	0.475
655	3.359*	3.959*
675	2.170*	0.707
679	7.042*	1.927
681	5.717*	5.717*
683	2.618*	2.728*
685	0.293	0.293
688	.	14.000*
697	3.103*	1.084
821	1.427	2.183*
842	2.183*	2.183*
894	2.499*	1.730
941	6.538*	0.531
951	1.855	2.824*
1994		
001	3.462*	2.729*
012	198.500*	1.280
023	3.116*	3.116*
025	0.473	0.454
044	.	1.100
047	2.000*	2.000*
251	3.279*	3.279*
264	0.222	0.222
265	5.853*	6.030*
271	0.286	1.738
283	2.196*	2.152*
532	0.595	0.421
613	2.486*	2.486*
671	2.949*	2.767*
687	2.119*	2.119*
688	.	0.750
896	0.450*	0.477
911	.	1225958.00*
931	.	0.208
961	1.222	34055.611*
999	.	.

(出所) 筆者作成

(注) *C,DW,UW* はそれぞれ商品分類コード、配分ウエイトおよび均等配分により得られたSITC-R1系列の対前年度比を示す。

し、最終調整においてeが0の対応関係を追加してすべての配分ウエイトを再計算している。

おわりに

本章はSITC-R1で編集された３桁レベルの分類コードもつ長期時系列データを作成することを目的とし、このための方法として商品グループごとに配分ウエイトを取引額にもとづいた制約条件付き最小２乗法で推計するという方法を採用した。配分ウエイトを推計するという方法を利用することによりSITC-R2あるいはSITC-R3の長期時系列も同じように作成が可能である。

この方法を実施していくにあたって、いくつかの課題が残されている。例えば、計算可能な商品グループの設定やサブグループを作成するための効果的な切断方法、商品グループが大きいときの推計方法などの課題がそうである。本章ではそれらの問題点は個別的に解決したが、この方法を日本の輸出以外に適用するにはより一般的なデータ処理方法が必要になる。

作成された長期時系列の評価が最も重要な検討課題である。比較の基準となる系列が存在しないため評価は難しいが、商品分類の改訂年の前後で断層が生じている分類コードは変換に問題がある可能性を含むと考えられる。野田の「商品分類の改定に伴う貿易統計の整合性評価」は断層の存在を商品分類の改訂時点と変化点が一致するかどうかで検証している。

本章では対前年比を利用して断層の存在を確かめている。表２に商品分類の改訂年である1987、88、94年のそれぞれについて対前年度比が比較的大きな3桁レベルの分類コードが示されている。この表において分類コードはC、推計された配分ウエイトにより作成された系列をDW、均等配分により作成された系列がUWである。対前年度比が2倍以上のものは*で示される。表では示していないが、年計の貿易統計データは商品分類コードによっては商品分類の改訂年前後にかかわらず対前年度比が極端に上下に変動することもある。したがって、表に示された分類コードのデータを利用する場合には十分な検討が必要とされる。

（注１）(1) 式と (2) 式はニューラルネットワークを利用して配分ウエイト行列Wを推計するさいの統計式として表現される。ニューラルネットワークによる配分ウエイトの推計方法は城坂による「SITC3桁分類コード変換のための配分ウエイト推計－ニューラル・ネットワークを用いて－」を参照すること。

（注２）対応関係のタイプ 2 では配分ウエイトは$\omega=(\omega_{11}\ \omega_{12}\cdots\omega_{1n})'$のみしか存在しない。調整済みの$\omega$で表現するための$D$は、$D_j=(1\ 0\cdots 0)'$, $j=1\cdots n$とするとき、

$$D=\begin{pmatrix} D_1 & & \\ & \ddots & \\ & & D_n \end{pmatrix}$$

となる。

$$\omega^D=D\omega=\begin{pmatrix} D_1 & & \\ & \ddots & \\ & & D_n \end{pmatrix}\begin{pmatrix}\omega_1\\ \vdots\\ \omega_n\end{pmatrix}=\begin{pmatrix}\omega_{11}\\ \vdots\\ \omega_{1n}\end{pmatrix}$$

$$C^D=CD'=(I_n\cdots I_n)\begin{pmatrix} D_1' & & \\ & \ddots & \\ & & D_n' \end{pmatrix}$$

$$=(D_1'\cdots D_n')=\begin{pmatrix}1&\cdots&1\\0&&0\\ \vdots&&\vdots\\0&\cdots&0\end{pmatrix}$$

この行列において 2 行目以降のすべての要素は 0 なので、2 行目からn-1 行分を削除して$C^D=(1\cdots 1)$となる。

（注３）アジア経済研究所が所有している OECD 貿易統計の日本データは 1962 年から 77 年までは商品分類が SITC-R1、78 年から 87 年までは SITC-R2、88 年から 93 年までは SITC-R3、94 年から 99 年までは HS-O、2000 年は HS-R1（1996 年度版）で編集されている。そのため、日本の 2000 年データは HS-R1 と SITC-R3 の対応表にしたがって SITC-R3 へと変換されるが、変換された SITC-R3 は HS-O と SITC-R3 の対応関係により変換された SITC-R3 とは桁レベルにおいて必ずしも分類コードが同一ではない。したがって、商品グループが 1999 年以前と 2000 年とでは異なるものが生ずることになる。本章では2000 年を対象から外してデータ処理を容易にしている。

（注４）貿易統計データの統計値の中で取引金額を対象としたとき、分類カテゴリーの商品分類および相手国についてはそれぞれの個別商品分類コードのみならずそれら合計値である商品総額および相手国世界がデータとして含まれている。そのため、商品総額と相手国世界のデータを基準値とすることによ

り個別商品分類コードあるいは個別相手国のデータを合計した値が基準値に一致するかどうかでそれぞれ商品分類および相手国の整合性を検討することができる。基準値から合計値を引いたものを誤差といい、このように個別商品分類コードあるいは個別相手国を合計することで誤差を評価する方法、すなわち、既存の総額と合計した値の比較をすることで分類カテゴリーの整合性を検討する方法をサムチェックという。整合性のあるとは思われない国については詳細分類における補正による調整をおこなっている。SITCの分類体系あるいはHS分類体系で編集されている貿易統計において下位桁レベルの分類コードが存在しているにもかかわらずその合計した取引額がその直接上位桁レベルの分類コードの取引額に一致しないものが存在することがある。一般的には商品分類の桁レベルにおける整合性の評価により上位桁レベル程整合性が高いことが知られている。したがって、下位桁レベルの分類コードを上位桁レベルにおいて合計した取引額とその上位桁レベルの分類コードの取引額の差が大きいときには下位桁レベルの分類コードを使わずにその上位桁レベルのそれを使用して整合性を高めることが必要になってくる。このような方法により整合性を高める処理を桁レベル分類コードの補正という。正確に言えばこの補正された貿易データがAID-XT基礎データである。

（注5）一般に配分のためのウエイト情報が事前情報として得られないとき等配分することで暫定的なウエイトを設定することが多い。等配分のウエイトは事前情報なしのエントロピーの最大化に根拠を置いており、ウエイトに一様分布が適用されるからである。対応関係の基本分類に基づく均等配分の方法は黒子の「商品分類の産業分類への変換－変換エラーデータの処理－」を参照すること。

（注6）図には示されていないが、その内訳は対応関係のタイプ1のものは369個、タイプ2が62個、タイプ3が169個、タイプ4aが219個、タイプ4bが115個となっている。

（注7）1つのグループからいくつかの対応関係を取り除くとグループがさらに2つ以上のグループに分かれるとき、この対応関係によってグループが「切断」されたといい、そのときに取り除いた対応関係を「切断の要素」ということにする。また、このとき得られたグループをもとのグループに対するサブグループという。グループを、

$$G = G_1 \cup \cdots \cup G_m$$

$G_i \cap G_j = \phi,\ (i \neq j)$ とする。グループの i 番目である G_i の切断の要素の集まりを C_i で表す。このグループから切断の要素を取り除いた $G_i - C_i$ に対してグループ化をおこないサブグループを作る。n_i 個のサブグループを $SG_i(j)\quad j = 1 \cdots n_i$ とするとグループ G_i は、

$$G_i = \{SG_i(1) \cup \cdots SG_i(n_i) \cup C_i\}$$

$SG_i \cap SG_j = \phi\ (i \neq j)$ と分割される。サブグループがグループから切断の要素を取り除いた対応関係コード表に対して再度グループ化をすることで得られるということは、切断の仕方によってサブグループの内容や個数が決まるということである。このことは切断というのは対応関係コード表のグループ化に対する1つのモデルであると考えることができる。切断をしない対応関係のモデルを対応関係の基本モデル、切断によりサブグループ化された対応関係を対応関係の切断モデルという。

【参考文献】

［1］黒子正人「商品分類の産業分類への変換－変換エラーデータの処理－」（野田容助編『世界貿易マトリクス－国際産業連関表24部門分類にもとづいて－』 統計資料シリーズ No.84 アジア経済研究所 2002）

［2］小西貞則「ブートストラップ法による推定量の誤差評価」（村上征勝、田村義保『パソコンによるデータ解析』朝倉書店 1988）

［3］――――「ブートストラップ法と信頼区間の構成」（『応用統計学』Vol.19 No.3,1990 応用統計学会 1991）

［4］城坂晃正「SITC3桁分類コード変換のための配分ウエイト推計－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］野田容助、山本泰子「体系の異なる分類の対応関係と変換－グループ化および切断による商品分類の変換の試み－」（木下宗七・野田容助編『世界貿易データシステムの整備と利用』統計資料シリーズ No.67 アジア経済研究所 1995）

［6］———「商品分類の改訂にともなう対応関係の連結」（古河俊一・野田容助共著『標準貿易商品分類と産業分類の対応関係』統計資料リーズ No.80 アジア経済研究所 1998）

［7］———「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』統計資料リーズ No.83 アジア経済研究所 2001）

［8］———「商品分類の改訂に伴う貿易統計の整合性評価」（野田容助編『商品分類の改訂に伴う貿易統計の変換』統計資料リーズ No.83 アジア経済研究所 2001）

［9］———「世界貿易マトリクスにおける整合性評価」（野田容助編『世界貿易マトリクス—国際産業連関表 24 部門分類にもとづいて—』統計資料リーズ No.84 アジア経済研究所 2002）

［10］United Nations, *Standard International Trade Classification, Revision 2*, Statistical Papers Series M no.34/Rev.2, New York, 1975

［11］United Nations, *Standard International Trade Classification, Revision 3*, Statistical Papers Series M no.34/Rev.3, New York, 1986

［12］United Nations, *Commodity Indexes for the Standard International Trade Classification, Revision 3*, Statistical Papers Series M No.38/Rev.2, Vol. II, New York, 1994