

第3章

世界貿易マトリクス作成における整合性の評価と補正

野田容助

はじめに

世界経済を構成する各国・各地域経済は国際依存のもとで世界経済のダイナミズムに組み込まれておりその関係はますます密接になってきている。国際比較のための世界貿易統計データとしてOECD作成の貿易統計（OECD貿易統計という）およびUN作成の貿易統計（UN貿易統計）が一般的に利用されている。この両作成機関の貿易統計データは主要な分類カテゴリーとして報告国、輸出入区分、商品分類、相手国、年、数量単位の6種類があり、統計値としては貿易取引数量と貿易取引金額の2種類から構成される。取引金額のみに興味がある場合には貿易統計は数量単位および数量を無視して5種類の分類カテゴリーと1種類の統計値によるデータ構成となる。

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。

本章は世界貿易マトリクスの対象国である東アジア諸国・地域および米国の11カ国について報告国、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評価をおこない、可能な限りその補正をおこなうことを目的としている。本章では世界貿易マトリクスの整合性の評価方法を紹介した後、特に、商品分類に関しては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おこなう方法を紹介する。世界貿易マトリクスにおける誤差評価のためのサムチェックにおいて統計値は貿易取引金額とそれに対応する取引数量の2種類が考えられる。しかし、数量単位ごとに処理しなければならない数量のデータ処理はかなり厄介であるため、本章ではその煩雑さを避けるため取引金額のみを整合性評価の対象とする。

1. 貿易マトリクスにおける整合性

貿易統計で使用される商品貿易分類は商品分類体系としてはUN作成の標準国際商品貿易分類（SITC: Standard International Trade Statistics）系列と関税協力理事会が作成する国際統一商品

表1 詳細なる商品分類コードmdccと個別相手国をもとに作成された世界貿易マトリクスの金額表

<i>C</i>	<i>P</i>	<i>P</i> ₁	<i>P</i> _{<i>j</i>}	<i>P</i> _{<i>n</i>}	<i>error of P</i>	<i>World</i>
<i>C</i> ₁		<i>x</i> ₁₁	<i>x</i> _{1<i>j</i>}	<i>x</i> _{1<i>n</i>}	<i>e</i> _{<i>p</i>} (1)	<i>x</i> _{1<i>W</i>}
:		:	:	:	:	:
<i>C</i> _{<i>i</i>}		<i>x</i> _{<i>i</i>1}	<i>x</i> _{<i>i</i><i>j</i>}	<i>x</i> _{<i>i</i><i>n</i>}	<i>e</i> _{<i>p</i>} (<i>i</i>)	<i>x</i> _{<i>i</i><i>W</i>}
:		:	:	:	:	:
<i>C</i> _{<i>m</i>}		<i>x</i> _{<i>m</i>1}	<i>x</i> _{<i>m</i><i>j</i>}	<i>x</i> _{<i>m</i><i>n</i>}	<i>e</i> _{<i>p</i>} (<i>m</i>)	<i>x</i> _{<i>m</i><i>W</i>}
<i>error of C</i>		<i>e</i> _{<i>c</i>} (1)	<i>e</i> _{<i>c</i>} (<i>j</i>)	<i>e</i> _{<i>c</i>} (<i>n</i>)	<i>e</i> _{<i>c,p</i>}	<i>e</i> _{<i>c</i>} (●)+ <i>e</i> _{<i>c,p</i>}
<i>Total</i>		<i>x</i> _{<i>T</i>1}	<i>x</i> _{<i>T</i><i>j</i>}	<i>x</i> _{<i>T</i><i>n</i>}	<i>e</i> _{<i>p</i>} (●)+ <i>e</i> _{<i>c,p</i>}	<i>x</i> _{<i>T</i><i>W</i>}

(出所) 筆者作成

(注) 影の部分は実際に得られるデータである。

分類あるいは統一システム (HS: Harmonized Commodity Description and Coding System) 系列が存在する。SITCの商品分類系列は商品総額のもとに1桁レベルから5桁レベルまでの各層に分けられた商品分類コードから構成されている。アジア経済研究所ではこの階層的に構成された商品分類コードの中で下位の階層の分類コードを持たないものを詳細分類コード (mdcc: most detail classification code) と呼んでいる^(注1)。例えば、SITC系列では分類の体系としては基本項目は4桁レベルの分類コードあるいは5桁レベルの分類コードから構成されているが、実際のデータにおいてはこれらの分類コードのみが必ずしも詳細分類コードであるとは限らない。3桁分類コードが詳細分類コードに含まれることもある。本章では商品分類における詳細分類コードの取引額をすべて合計すると商品総額に一致することを整合性の評価基準とする。

詳細分類コードから構成される商品分類を $C = \{C_1 \dots C_m\}$ 、個別相手国から構成されている相手国を $P = \{P_1 \dots P_n\}$ とするとき、貿易統計として得られる取引金額のデータは商品分類 C_i と相手国 P_j に対して報告国 (reporting country: *rc*)、年 (year: *y*)、輸出入区分 (flow of

goods, type of trade (data): *d*) ごとに、

$$(1) \quad x_{ij}(rc, d, y) \quad i \in C, j \in P$$

と表わすことができる。報告国、年、輸出入区分を固定すれば簡単に $x_{i,j}$ と表わされる。前述したように実際の貿易統計データには商品分類に関しては商品総額 (*Total*) が含まれているため、正確には商品分類は $C^* = \{Total, C\}$ であり、相手国についても相手国合計としての世界 (*World*) が含まれているため、 $P^* = \{World, P\}$ となるのが一般的である。

得られたデータを表頭に相手国、表側に商品分類を対応させて作成した貿易マトリクスが表1に示されている。実際にデータとして得られるものは影を付け x_{ij} と表示している。商品分類の *i* について商品総額の *Total* を *T* で表わし、相手国の *j* については世界の *World* を *W* で表わすことにすれば、 x_{TW} は商品総額であり同時に相手国世界の取引金額となる。この x_{TW} を基準とすることにより貿易マトリクスの整合性の評価が可能となる。後述するようにこの貿易マトリクスには整合性を保つためには相手国および商品分類についてそれぞれの誤差の項目である *error of P* (Partner countriesに関する誤差項目) と *error of C* (Commoditiesに関する誤差項目) が必要と

なるが、それらのデータをそれぞれ e_p と e_c で表わすことにする。

貿易統計データをもとにして作成された貿易マトリクスの取引金額において、商品分類を C_i に固定して相手国を合計しても必ずしも相手国世界に一致するとは限らない。一般的には得られたデータに関して相手国 j について合計して、

$$x_{i\bullet} = \sum_{j=1}^n x_{ij} \quad i=1\cdots m$$

とすると、 $x_{iW} \neq x_{i\bullet}$ である。この原因としては丸め誤差や相手国不明の取り扱いによる誤差が考えられる。商品分類 C_i の相手国世界 x_{iW} を基準としてこの誤差を相手国による誤差 $e_p(i)$ とすると、

$$(2) \quad e_p(i) = x_{iW} - x_{i\bullet} \quad i=1\cdots m$$

と表わされる。表1の貿易マトリクスにおいて相手国に関する誤差は*error of P*の項目に対応する。同じことが商品分類についてもいえる。相手国を P_j と固定したとき、商品分類を合計しても必ずしも商品総額に一致しない。

$$x_{\bullet j} = \sum_{i=1}^m x_{ij} \quad j=1\cdots n$$

とすると、 $x_{Tj} \neq x_{\bullet j}$ である。商品分類の総額を基準としてこの誤差を商品分類による誤差 $e_c(j)$ とすると、

$$(3) \quad e_c(j) = x_{Tj} - x_{\bullet j} \quad j=1\cdots n$$

となる。表1の貿易マトリクスにおいて商品分類に関する誤差は*error of C*の項目に対応する。さらに、表1における*error of P*の項目と*error of C*の項目の交点を相手国および商品分類の共通の誤差として $e_{c,p}$ とする。以上により詳細分類にもとづく整合性のある世界貿易マトリクスは完成する。

次に誤差の推定値を得るために得られたデータの相手国および商品分類の合計を、

$$x_{\bullet\bullet} = \sum_{i=1}^m \sum_{j=1}^n x_{ij} \quad i=1\cdots m, \quad j=1\cdots n$$

とする。(2)式において商品分類の合計を、

$$(4) \quad e_p(\bullet) = \sum_{i=1}^m e_p(i) = \sum_{i=1}^m (x_{iW} - x_{i\bullet}) \\ = x_{\bullet W} - x_{\bullet\bullet} = e_p$$

とすると、この誤差は商品分類とは無関係な相手国のみの誤差となる。また、(3)式における相手国の合計を、

$$(5) \quad e_c(\bullet) = \sum_{j=1}^n (x_{Tj} - x_{\bullet j}) = x_{T\bullet} - x_{\bullet\bullet} = e_c$$

とすると、この誤差は相手国とは無関係な商品分類のみによる誤差となる。相手国および商品分類の共通の誤差は、

$$e_{c,p} = x_{TW} - e_p(\bullet) - \sum_{j=1}^n x_{Tj} \\ (6) \quad = x_{TW} - e_c(\bullet) - \sum_{j=1}^n x_{iW} \\ = (x_{TW} + x_{\bullet\bullet}) - (x_{T\bullet} + x_{\bullet W})$$

と表わされる。(4)式から(6)式までにより、貿易マトリクスの総合誤差を e とすれば、

$$(7) \quad e = x_{TW} - x_{\bullet\bullet} = e_c + e_p + e_{c,p}$$

が得られる。

本書では報告国、年、輸出入区分を固定したときの詳細なる商品分類と個別相手国をもとにした貿易マトリクスの誤差表示を総合誤差 e 、商品分類のみから生じた誤差を(5)式と(6)式の和として、

$$e_c + e_{c,p} = x_{TW} - x_{\bullet W}$$

と表し、相手国のみから生じた誤差を(4)式と(6)式の和として

$$e_p + e_{c,p} = x_{TW} - x_{T\bullet}$$

として表わすことにする(註2)。

2. 商品分類における整合性評価

貿易統計データには商品分類に関して前述したようにそれぞれ桁レベルにおける取引額のデータが存在する(註3)。したがって、上位桁レベ

表2 商品分類の k 桁レベル合計と相手国合計をもとに作成された世界貿易マトリクスの金額要約表

C	P	P : 個別相手国の合計	$Error\ of\ P$	$World$
$C[k]$: 商品分類の合計		$x_{\bullet\bullet}[k]$	$e_p[k]$	$x_{\bullet W}[k]$
$error\ of\ C[k]$		$e_c[k]$	$e_{c,p}[k]$	$e_c[k]+e_{c,p}[k]$
$Total$		$x_{T\bullet}[k]$	$e_p[k]+e_{c,p}[k]$	x_{TW}

(出所) 筆者作成

表3 商品分類SITC系列およびHS系列の各改訂版における桁レベル分類コードの名称とその項目数

(1) SITC の系列

分類のレベル 改訂版の種類	大分類 (1 桁)	中分類 (2 桁)	小分類 (3 桁)	細分類 (4 桁)	基本項目 (4,5 桁)
SITC-R1	10	56	177	625	1,312 (368)
SITC-R2	10	63	233	786	1,832 (365)
SITC-R3	10	67	261	1,033	3,121 (299)

(2) HS の系列

分類のレベル 改訂版の種類	部 (2 桁)	類 (2 桁)	項 (4 桁)	号 (6 桁)
HS88	21	97	177	5,019
HS96	21	97	233	5,114

(出所) 野田容助「商品分類における詳細分類コードの抽出」(『世界貿易マトリクスの作成と評価—貿易指数の推計に向けて』調査研究報告書 アジア経済研究所 2002)の表1および表3にもとづき筆者作成。

(注) SITC系列の基本項目において表示されている数字は4桁レベル分類コードおよび5桁レベル分類コードの項目数合計であり、()の中の数字は4桁レベル分類コードのみの項目数を表す。

ルを規準とするとき、その基準値と下位の桁レベルで表示されているものを合計した値と比較することにより桁レベルの整合性も検討できる。商品分類の k 桁レベル分類コードで表された取引額を(1)式と同じように表すことができ、 $k=0,1,\dots,4,mdcc$ に対して、

$$x_{ij}(rc,d,y)[k] \quad i \in C, j \in P$$

として、簡単に $x_{ij}[k]$ と表すことにする。サムチェックの定義により商品総額るときには $k=0$ であり、 $x_{Tj} = x_{Tj}[0] = x_{\bullet j}[0] = x_{ij}[0]$ となる。それぞれの誤差についても同じように、

$e_p[k], e_c[k], e_{c,p}[k]$ となる。商品総額に対して誤差は生じないので、 $e_p[0], e_c[0], e_{c,p}[0] = 0$ である。詳細分類コードのときには k は $mdcc$ となる。表1の商品別相手国別に表示された世界貿易マトリクスでは桁レベルを省略しているが、商品分類は詳細分類コード $mdcc$ の表示である。

商品分類の k 桁レベル合計と相手国合計をもとに作成された取引金額要約表が表2に示される。この桁レベルにおけるサムチェックは詳細分類において不整合が生じたときの次のステップとしての整合性の検討である。商品総額は桁

の最高レベルであるので上位桁レベルを商品総額に固定することも可能である。商品分類における整合性の評価として以下、詳細分類コードによる整合性、桁レベルにおける整合性、桁レベルの分類コードに対する整合性を説明する。

表3に示されているように貿易統計ではSITCの分類体系において基本項目である詳細分類コードは4桁あるいは5桁レベルの分類コードで定義されている。しかし、実際のデータにあつては必ずしもそうであるとは限らない。例えば、HSとSITC-R3の対応関係においてHSの271000はSITC-R3の334にしか対応していないために国によっては334以下の分類コードを割当てていこともあるがこの対応関係を忠実に実行している国では3桁が詳細分類コードになっている^(注4)。したがって、詳細分類コードを抽出しようとするときにはすべての桁レベルの分類コードが抽出の対象となる。

また、SITCの分類体系で編集されているUNあるいはOECD貿易統計において5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が4桁レベルに一致しないものが存在する。そのため詳細分類コードを合計しても商品総額に一致しないという状態がおこる。このように詳細分類コードに不整合があると見なされたとき、どの桁レベルが商品総額に一致しており利用に値するかを判断するのが商品分類の桁レベルによる整合性の評価である。(1)式において報告国、輸出入区分、年にさらに相手国世界を固定して $i \in C$ に対して x_{iW} を利用する。商品分類がSITCの分類体系ではそれぞれの桁レベルは商品合計を0桁レベル、大分類を1桁レベル、中分類を2桁レベル、小分類を3桁レベル、細分類を4桁レベルとしているので、桁レベルの合計を $k=0, \dots, 4$ に対して $x_{iW}[k]$ として表す。もちろん商品総額は $x_{iW}[0]=x_{TW}$ である。サムチェックは商品総額が基準となるのでそれぞれの桁

レベルに対する評価として、 $k=1 \dots 4$ に対して絶対誤差を、

$$(8) \quad a_k = x_{TW} - x_{iW}[k]$$

とし、その相対誤差を

$$(9) \quad b_k = (x_{TW} - x_{iW}[k]) / x_{TW}$$

する。また、詳細分類コードに対する k は $mdcc$ に相当する。

商品分類がHSである分類体系では表3により、類(Chapter)を2桁レベル、項(Heading)を4桁レベル、号(Sub-heading)を6桁レベルとしているほか、Chapterを分割して得られる2桁の最上位レベルの部(Section)が存在する。それぞれの桁レベルにおける取引額の合計を $k=0, 2, 4, 6$ に対して $x_{iW}[k]$ と表す。ただし、Sectionについては上位レベルの概念が他のそれとは異なっており煩わしいので本章では考察の対象とはしないことにする。サムチェックのでそれぞれの桁レベルに対する評価の絶対誤差および相対誤差は $k=0, 2, 4, 6$ に対して(8)式と(9)式で与えられる。

桁レベルの合計で整合性に問題があると判断されたとき、さらなる検査として桁レベルの分類コードにおける誤差の検査が必要になる。商品分類がSITC分類体系のとき、(1)式とは異なる表現を用いて取引額を、

$$(11) \quad x(rc, d, y, W; i_1, i_2, i_3, i_4, i_5)$$

と表す。 $k=1 \dots 5$ に対して、 $i_k=0 \dots 9$ である。明示的には示していないが、分類コードによっては $i_k=-$ を含むこともある。(11)式を省略して、 $x(i_1, i_2, i_3, i_4, i_5)$ とする。また、(11)式は k 桁レベルの分類コード $\{i_1, \dots, i_k\}$ の取引額を $x(i_1, \dots, i_k)$ と表し、 $k+1$ 桁レベルの分類コードを合計して得られた k 桁レベルの分類コード $\{i_1, \dots, i_k\}$ の取引額を、

$$x(i_1, \dots, i_k, \bullet) = \sum_{i_{k+1}=0}^9 x(i_1, \dots, i_k, i_{k+1})$$

と表しその違いがわかるようにしている。明示

的には示していないが、分類コードによっては合計のさいに $i_{k+1} = -$ を含むこともある。

商品分類がHS分類体系のときには、(11)式に対応して、

$$(12) \quad x(rc, d, y, W : i_2, i_4, i_6)$$

とする。 $k=2,4,6$ に対して、 $i_k = 00 \dots 99$ である。(12)式を省略して、 $x(i_2, i_4, i_6)$ と表わす。ここでも、Section1については上位レベルの概念が他のそれとは異なっており煩わしいので省略する。

世界貿易マトリクス作成におけるmdccレベル分類コードの整合性評価表は報告国、輸出入区分、年毎に、 x_{TW} 、総合誤差 $e[mdcc]$ 、各桁レベル分類コードの個数 $\{d_1 \dots d_6\}$ 、商品分類による誤差 $e_c[mdcc] + e_{c,p}[mdcc]$ 、相手国による誤差 $e_p[mdcc] + e_{c,p}[mdcc]$ のそれぞれ符号付き絶対誤差と x_{TW} に対するその誤差の割合を順に並べて表示する。整合性評価表の例は表4に示されている。

3. 詳細分類コードにおける補正

商品分類における整合性の評価は以上の通りであるが、5桁レベル分類コードとその上位の階層である4桁レベル分類コードの間ではデータについての構造上の問題が生じている。SITCの分類体系で編集されているUNあるいはOECD貿易統計において5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4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である。一般に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、5桁レベル分類コードを4桁レベルにおいて合計した取引額と4桁レベル分類コードの取引額との差が大きいときには5桁レベルの分類コードを使わずにその上位の桁レベルであ

る4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4桁レベル分類コードに対する補正という。

補正は4桁レベル分類コードのデータと4桁レベルへ集計された5桁レベル分類コードのデータに対する絶対誤差と相対誤差の2種類によって補正基準が決められる。絶対誤差による補正は絶対誤差を a とするとき(11)式を利用すれば、

$$\alpha_4 = |x(i_1, i_2, i_3, i_4) - x(i_1, i_2, i_3, i_4, \bullet)|$$

となり、その補正基準を α_4^* とすれば、補正は、

$$(13) \quad \alpha_4 \geq \alpha_4^*$$

としたとき、(13)式を満たす4桁レベルの分類コード $\{i_1, \dots, i_4\}$ は5桁レベルの商品分類が存在していてもその4桁レベルを分類コードとして採用することである。(13)式を満たさないときは5桁レベル分類コードを採用する。ここで、 α_4^* を4桁レベルにおける絶対補正係数といい、 $\alpha_4^* \geq 0$ である。

また、相対誤差による補正基準は相対誤差 b を、

$$\beta_4 = \frac{|x(i_1, i_2, i_3, i_4) - x(i_1, i_2, i_3, i_4, \bullet)|}{x(i_1, i_2, i_3, i_4)}$$

として、その補正基準を β_4^* とすれば、

$$(14) \quad \beta_4 \geq \beta_4^*$$

となるとき、(14)式を満たす4桁レベルの分類コード $\{i_1, \dots, i_4\}$ は5桁レベルの商品分類が存在していても4桁レベルを分類コードとして採用することである。ここで、 β_4^* を相対補正係数といい、 $0 \leq \beta_4^* \leq 1$ である。

補正係数が $\alpha_4^* = 0$ のときは β_4^* も0となり、すべての5桁レベルの分類コードを4桁レベルのそれへ置き換えることを意味する。一方では、 $\beta_4^* = 1$ のときは丸めの誤差により必ずしも正しくはないが、一般に貿易統計では、

$$x(i_1, i_2, i_3, i_4) - x(i_1, i_2, i_3, i_4, \bullet) > 0$$

であるので(14)式において、 $x(i_1, i_2, i_3, i_4) < 0$

表4 OECD 貿易統計における日本の1994-2001の詳細分類コードによる取引金額の整合性評価

y	x_{TW}	e	d_1	d_2	d_3	d_4	d_5	d_6	$e_c + e_{c,p}$	$e_p + e_{c,p}$			
(import)													
1994	276130918	-2831	-0.00001	0	0	0	0	0	4860	-215	0.00000	-7	0.00000
1995	336141246	-3040	-0.00001	0	0	0	0	0	4861	-218	0.00000	-8	0.00000
1996	349186136	6648211	0.01904	0	0	0	0	0	4776	6651126	0.01900	-8	0.00000
1997	338812068	6065833	0.01790	0	0	0	0	0	4781	6069051	0.01700	-8	0.00000
1998	280633885	4157125	0.01481	0	0	0	0	0	4762	4160499	0.01400	-8	0.00000
1999	309915152	5485332	0.01770	0	0	0	0	0	4770	5488272	0.01700	-12	0.00000
2000	379667028	6862818	0.01808	0	0	0	0	0	4779	6865752	0.01800	-10	0.00000
2001	348604783	-3550	-0.00001	0	0	0	0	0	4988	-243	0.00000	-6	0.00000
(export)													
1994	397731730	-5746	-0.00001	0	0	0	0	0	4606	-240	0.00000	-5	0.00000
1995	443251353	-5835	-0.00001	0	0	0	0	0	4614	-223	0.00000	-9	0.00000
1996	410947004	7564850	0.01841	0	0	0	0	0	4518	7570448	0.01800	-13	0.00000
1997	421012181	9319237	0.02214	0	0	0	0	0	4542	9325260	0.02200	-13	0.00000
1998	388136220	8617584	0.02220	0	0	0	0	0	4504	8624203	0.02200	-9	0.00000
1999	417138039	8581336	0.02057	0	0	0	0	0	4531	8586944	0.02000	-11	0.00000
2000	479223369	10780095	0.02249	0	0	0	0	0	4550	10785527	0.02200	-10	0.00000
2001	402604490	-6331	-0.00002	0	0	0	0	0	4702	-243	0.00000	-19	0.00000

(出所) OECD 貿易統計にもとづき筆者作成

(注) yは西暦年、 x_{TW} は相手国世界における商品総額の取引額、単位は1,000US\$である。 d_k はk桁レベル分類コードの個数、eは総合誤差、 $e_c + e_{c,p}$ は商品分類にかかわる誤差、 $e_p + e_{c,p}$ は相手国にかかわる誤差を表し、それぞれの誤差の後ろの数字は x_{TW} に対する割合である。商品分類は1994年から2000年まではHS-O、2001年はHS-R1で編集されている。

となる。このようなことは起こり得ないので5桁レベルの4桁レベルへの置き換えはないことになる。このようにして α_4^* および β_4^* の値を適当に選択することにより5桁レベルの調整をする場合と同じように調整をしない場合にも利用できる。

商品分類がHSの分類体系でも同じようなことが生ずる。6桁レベル分類コードで表される取引金額を(12)式を利用して $x(i_2, i_4, i_6)$ とするとき、4桁レベルへ集計された6桁レベル分類コードの取引金額は $x(i_2, i_4, \bullet)$ と表される。この2つのデータを比較して生ずる絶対誤差を(13)式に対応させるとその誤差は、

$$\alpha_4 = |x(i_2, i_4) - x(i_2, i_4, \bullet)|$$

となり、その補正基準を α_4^* とすれば、

$$(15) \quad \alpha_4 \geq \alpha_4^*$$

となる時、補正は(15)式を満たす4桁レベルの分類コード $\{\dots, i_4\}$ は6桁レベルの商品分類が存在していてもその4桁レベルを分類コードとして採用することである。(13)式を満たさないうときは6桁分類コードを採用する。また、相対誤差による補正基準は相対誤差を、

$$\beta_4 = \frac{|x(i_2, i_4) - x(i_2, i_4, \bullet)|}{x(i_2, i_4)}$$

として、その補正基準を β_4^* とすると

$$(16) \quad \beta_4 \geq \beta_4^*$$

において、(16)式を満たす4桁レベルの分類コード $\{i_1, \dots, i_4\}$ は6桁レベルの商品分類が存在していても4桁レベルを分類コードとして採用することである。

表4にOECD 貿易統計の報告国日本における1994年から2001年までの詳細分類コードの整合

表5 OECD 貿易統計の日本の1994-2001の桁レベルによる総合誤差の整合性の評価

y	x_{TW}	e[2]	e[4]	e[6]	y	x_{TW}	e[2]	e[4]	e[6]
(import)					(export)				
1994	276130918	-282	-1226	-2831	1994	397731730	-365	-2259	-5746
1995	336141246	-296	-1355	-3040	1995	443251353	-376	-2343	-5835
1996	349186136	-320	-1348	6648211	1996	410947004	-408	-2466	7564850
1997	338812068	-338	-1444	6065833	1997	421012181	-444	-2225	9319237
1998	280633885	-300	-1460	4157125	1998	388136220	-444	-2727	8617584
1999	309915152	-296	-1358	5485332	1999	417138039	-384	-1835	8581336
2000	379667028	-275	-1442	6862818	2000	479223369	-357	-1805	10780095
2001	348604783	-324	-1564	-3550	2001	402604490	-421	-2431	-6331

(出所) OECD 貿易統計にもとづき筆者作成

(注) yは西暦年、 x_{TW} は相手国世界における商品総額の取引額、単位は1,000US\$である。e[2]は2桁レベルによる総合誤差、同じく、e[4]、e[6]はそれぞれ4桁レベルおよび6桁レベルにおける総合誤差を表す。輸出入ともに商品分類は1994年から2000年まではHS-O、2001年はHS-R1で編集されている。表4はすべての分類コードが6桁レベルであるのでその総合誤差と本表のe[6]とは一致している。

表6 OECD 貿易統計の日本の1999年輸入額における誤差を含む4桁レベル分類コード

C	$x(C)$	$x(C, \bullet)$	a	b	C	$x(C)$	$x(C, \bullet)$	a	b
0207	1000116	965288	34828	0.03482	7211	13369	13278	91	0.00681
0901	863368	862923	445	0.00052	7212	9572	8608	964	0.10071
2503	176	.	.	.	7213	12206	.	.	.
2513	4147	1810	2337	0.56354	7214	6842	5626	1216	0.17773
2848	3233	.	.	.	7215	1267	99	1168	0.92186
2914	70380	59672	10708	0.15215	7414	1339	1110	229	0.17102
3002	570255	544227	26028	0.04564	7907	9612	.	.	.
4403	2332695	2105000	227695	0.09761	8005	1007	.	.	.
4407	2995246	2855658	139588	0.04660	8207	168406	150734	17672	0.10494
4412	1933602	986143	947459	0.49000	8469	48974	48370	604	0.01233
4706	59051	57558	1493	0.02528	8471	12511350	10478479	2032871	0.16248
4807	1117	3	1114	0.99731	8479	1535614	1502808	32806	0.02136
5402	123222	119663	3559	0.02888	8517	3356724	2422548	934176	0.27830
6305	118138	42622	75516	0.63922	8520	90728	89708	1020	0.01124
6402	936562	860485	76077	0.08123	8524	792063	786901	5162	0.00652
6403	776815	766979	9836	0.01266	8528	1410352	1383229	27123	0.01923
7208	765757	6	765751	0.99999	9007	17088	7237	9851	0.57649
7210	188097	142588	45509	0.24194	9030	733813	704525	29288	0.03991

(出所) OECD 貿易統計にもとづき筆者作成

(注) 絶対誤差の補正基準 α_4 *は80と設定されている。CはHSの商品分類体系における4桁レベルの商品分類コード $\{i_2, i_4\}$ 、 $x(C) = x(i_2, i_4)$ は4桁レベル分類コードの取引額を表し、 $x(C, \bullet) = x(i_2, i_4, \bullet)$ は6桁レベル分類コードの取引額を4桁レベルへ集計した取引額を表す。単位はともに1,000US\$である。「。」は欠損値を表す。aは総合誤差の絶対誤差、bは同じく相対誤差を示す。

性評価表が示されている。この評価表の意味することは、輸入は1994年と1995年、2001年において総合誤差 e は絶対誤差においてそれぞれ4桁の-2,831、-3,040、-3,550であり、その相対誤差において0.00001あるいは0.00002である。商品分類にかかわる誤差および相手国にかかわる誤差は絶対誤差および相対誤差においてともに小さい。輸出についても同じような状況である。これらの誤差は個別相手国および詳細分類にもとづく個別商品分類コードを合計したときに生ずる丸めの誤差の範囲と解釈される。ところが、1996年から2000年の間では輸出入ともに相手国にかかわる誤差は丸めの誤差の範囲とみなすことができるが、商品分類にかかわる誤差については絶対誤差にして7,8桁、それは相対誤差では約2%に相当する大きな割合を示している。このことは商品分類については検討すべき内容が含まれていると解釈される。なお、OECD貿易統計の日本において1994年から2000年までは商品分類はHS-O、2001年はHS-R1で編集されており、2001年に整合性が保証されているのは商品分類の違いによると思われる。

表4の整合性評価表で示された商品分類の不整合の状態を桁レベルの総合誤差で示したのが表5である。表4で示された総合誤差の大きいのは1996年から2000年までであるが、これらの年は表5により輸出入ともに4桁レベルの分類においては整合性が保証されていることがわかる。したがって、この場合には4桁レベルに対する補正が可能となる。

補正の対象となる商品分類の不整合状態にある4桁分類コードを示したのが表6である。表6はOECD貿易統計における日本の1999年輸入額であり絶対誤差の補正基準 α_4^* を80としたときの誤差を含む商品分類を4桁レベル分類コードの取引額およびその下位レベル分類の6桁レベル分類コードの合計額を共に表示したもので

ある。誤差を含む4桁レベル分類コードの個数は36個である。例えば、HSの4桁レベル分類コード0207は取引額が1,000,116 (1,000US\$)であり、6桁レベル分類コードの合計額は965,288であるため、総合誤差の絶対誤差 a が34,828、その相対誤差 b が0.03482のように示されている。6桁レベルの分類コードが存在せずに4桁レベルの分類コードのみが存在する分類コードは6桁レベル分類コードが欠損値の、で表されており、2503、2848、7213、7907、8005の5個存在する。

表6で示された4桁レベル分類コードの6桁レベル分類コードによる内訳を示したのが表7である。この表は表6において相対誤差 b が0.20以上となる分類コードの2513、4412、6305、7210、7215、8517、9007の7個のみを表示している。表6の商品分類コード2513において6桁レベル分類コードの合計2513..の取引額は1,810であり、絶対誤差は2,337と示されている。表7では2513と2513..と同時にその内訳である6桁レベル分類コードとその取引額が示され、251311は276、251319は1,534であり、その合計として2513..は1,810であることを示している。

さらに、表6で示された商品分類の不整合の補正係数を与えて調整した詳細分類による整合性の評価が表8に示されている。この表からわかるように、1994年、1995年および2001年については4桁レベル分類コードへの置き換えの必要がないため、総合誤差 e には変化が見られない。1996年から2000年までについては6桁レベルの分類コードから4桁レベルの分類コードの置き換えがおこなわれている。表8によりこの置き換えにより、1996年から2000年までの総合誤差と商品分類による誤差ともに丸めの誤差の範囲に調整されたことがわかる。

置き換えにより表4では0であった4桁レベル分類コードの個数 d_4 が増加していることがわかる。前述したように、表6において1999年輸入

表7 日本の1999年の輸入における相対誤差が20%以上である商品分類コードとその取引額

C	取引額	C	取引額	C	取引額
2513	4147	630520	230	7215..	99
2513..	1810	630531	33601	721510	24
251311	276	630539	4287	721590	75
251319	1534	630590	240		
				8517	3356724
4412	1933602	7210	188097	8517..	2422548
4412..	986143	7210..	142588	851710	171649
441212	737977	721012	19102	851720	14
441219	120034	721020	13	851730	885664
441221	177	721041	40	851790	1365221
441229	72693	721049	90798		
441291	45	721050	8159	9007	17088
441299	55217	721060	12601	9007..	7237
		721070	7989	900711	202
6305	118138	721090	3886	900719	2290
6305..	42622	7215	1267	900791	2093
630510	4264			900792	2652

(出所) 表4に同じ

(注) 相手国世界における商品分類コードとその取引額を示し、単位は1,000US\$である。Cは商品分類コードを表し、2513の取引額は $x(25,13)$ であり、2513..の取引額は $x(25,13,\bullet)$ である。2513..以下に続く6桁レベル分類コードはその内訳である。

の誤差を含む4桁レベル分類コードの個数は36個であり、その中に6桁レベル分類コードの欠損値の5個があるため、これを差し引いた31個が実質の置き換えの個数である。欠損値の5個は4桁レベル分類コードが追加される。表8から置き換えによる調整後の d_4 が示され、1999年の輸入は36個であることが確認できる。

補正係数の決め方は容易ではない。例えば貿易指数を作成するときにはできるだけ細かい分類を使用して数量データを利用することが必要となる。このような場合には、できるだけ細かい分類の方が望ましいという立場に立って5桁レベルの分類コードの存在を考慮しつつ整合性を高めるように補正係数を決めることになる。しかし、分類の細かさより整合性のほうが重要であることもあるかもしれない。補正係

数の決定方法はまだ確定しておらず今後の課題として残されている。

おわりに

本章で取り上げたのは世界貿易マトリクス作成において最初に問題となる貿易統計データの整合性についての検討であり、対象としている商品分類に対する貿易取引額が商品総額に一致することを基準とするサムチェックによる評価方法である。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整合性のある詳細分類コードによる

表8 OECD 貿易統計における調整された日本の1996-2000の取引金額の整合性評価

year	x_{TW}	e	d_1	d_2	d_3	d_4	d_5	d_6	$e_c + e_{c,p}$	$e_p + e_{c,p}$			
(import)													
1994	276130918	-2831	-0.00001	0	0	0	0	0	4860	-215	0.00000	-7	0.00000
1995	336141246	-3040	-0.00001	0	0	0	0	0	4861	-218	0.00000	-8	0.00000
1996	349186136	-3012	-0.00001	0	0	0	37	0	4605	-120	0.00000	-8	0.00000
1997	338812068	-3161	-0.00001	0	0	0	37	0	4616	0	0.00000	-8	0.00000
1998	280633885	-3440	-0.00001	0	0	0	37	0	4598	-134	0.00000	-8	0.00000
1999	309915152	-2999	-0.00001	0	0	0	36	0	4608	-146	0.00000	-12	0.00000
2000	379667028	-3049	-0.00001	0	0	0	36	0	4619	-161	0.00000	-10	0.00000
2001	348604783	-3550	-0.00001	0	0	0	0	0	4988	-243	0.00000	-6	0.00000
(export)													
1994	397731730	-5746	-0.00001	0	0	0	0	0	4606	-240	0.00000	-5	0.00000
1995	443251353	-5835	-0.00001	0	0	0	0	0	4614	-223	0.00000	-9	0.00000
1996	410947004	-5564	-0.00001	0	0	0	33	0	4385	-101	0.00000	-13	0.00000
1997	421012181	-5635	-0.00001	0	0	0	33	0	4410	252	0.00000	-13	0.00000
1998	388136220	-6511	-0.00002	0	0	0	32	0	4381	-9	0.00000	-9	0.00000
1999	417138039	-5082	-0.00001	0	0	0	33	0	4404	418	0.00000	-11	0.00000
2000	479223369	-4727	-0.00001	0	0	0	29	0	4433	619	0.00000	-10	0.00000
2001	402604490	-6331	-0.00002	0	0	0	0	0	4702	-243	0.00000	-19	0.00000

(出所) 表4に同じ

(注) 表4に同じ

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。そのため、AID-XT基礎データの評価方法は報告国、輸出入区分ごとに相手国世界である商品総額を比較基準としたサムチェックによる誤差評価が中心となる。

AID-XT基礎データ作成のデータ処理過程において、 $x_{..}, x_{T.} = 0$ であり、総合誤差 e と相手国による誤差は $e_p + e_{c,p}$ はともに変化せず x_{TW} であるが、商品による誤差は $e_c + e_{c,p} = x_{TW} - x_{.W}$ と表わされる。また、商品総額のみに限ったときは、 $x_{..}, x_{.W} = 0$ であり、総合誤差 e と商品による誤差は $e_c + e_{c,p}$ はともに変化せず x_{TW} であるが、相手国による誤差は $e_p + e_{c,p} = x_{TW} - x_{T.}$ と表わされる。

本書の目的である貿易マトリクス作成における貿易データの整合性の評価として東アジア諸国・地域および米国の11カ国に対してまとめた

のが第2部の表1である。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。しかし、補正の方法について絶対誤差あるいは相対誤差を用いて年、輸出入区分を考慮せずに報告国ごとに一括してそれぞれの基準を決めており、しかも場合によってはかなり恣意的に決めざるを得ないこともあった^(注5)。こうした補正の方法について相対的な基準のもとに輸出入、年ごとに個別に調整できるような方法論の改善が必要とされる。貿易統計の整合性をどの程度まで維持する必要があるかという問題も含め補正の方法については今後に残されている課題は多い。

(注1) 野田容助「商品分類における詳細分類コードの抽出」に詳細分類コードについての説明があ

るので紹介する。商品分類コードのなかで最も細かい分類コードである基本項目(アイテム:itemという)は概念的にはSITCの各改訂版では4桁あるいは5桁レベル、HSでは6桁のSub-Headingで定義されている。実際のデータでは詳細分類コードは必ずしも基本項目のみでは表現されない。しかも、詳細分類コードは報告国、輸出入区分および年によって異なるので、データの抽出には報告国、輸出入区分および年ごとの処理が必要となる。商品分類のSITCおよびHSについてそれぞれの分類体系に応よらない抽出方法を示す。詳細分類コードの桁数を m とする。ここで、 $m=1,2,\dots,6$ とする。(1) 分類コードの m 桁目が空白でないものを取り出す、(2) 分類コードの m 桁目までが同一であるものをグループとみなし、そのグループの先頭に1を付け、2番目以降に0を付ける、(3) さらに、そのグループに属する分類コードの個数をそれぞれの分類コードに付ける、(4) 各分類コードに付けられた2つの数字がともに1であるものの中から分類コードの桁数が m であるものだけを取り出す、(5) この処理を m が1から6まで繰り返す、(6) 得られた分類コードをまとめ、これを昇順に並べる。このようにして得られた分類コードの集まりが詳細分類コードである。繰り返しになるが、詳細分類コードは報告国、輸出入区分および年によって異なるので、そのための処理は報告国、輸出入区分および年ごとにおこなわなければならない。商品分類の詳細分類コード抽出のための外部サブルーチンが@DetC.pliである。パラメータは2つあり、すべての桁レベルの分類コードを含む文字属性を持つ1次元の配列変数 w とその分類コードの個数 n である。結果である詳細分類コードその個数はそれぞれ入力のパラメータと同一の w と n に戻ってくるので、必要ならば入力の変数は別の変数に記憶させておかなければならない。なお、@DetC.pliには詳細分類コードの空白部()を挿入する機能は含まれていない。

(注2) 貿易統計データにおいて整合性のある状態では誤差は存在せず $e_c = 0, e_p = 0, e_{c,p} = 0$ である。貿易統計データが整合性に欠けるときは以下の通り5つの誤差のタイプが考えられる。(1) 商品分類 C_i における相手国 P_j に誤差があるとして整合性のある貿易統計データに対して Δx_{ij} 増加した誤差である。

誤差は $e_c = -\Delta x_{ij}, e_p = -\Delta x_{ij}, e_{c,p} = \Delta x_{ij}$ となるのに対して、 $e_c + e_{c,p} = 0$ と $e_p + e_{c,p} = 0$ なり、これらの誤差は表示されない。総合誤差 e は $-\Delta x_{ij}$ となる。

(2) 商品分類 C_i の相手国世界が Δx_{iW} 増えたとなると、 $e_c = 0, e_p = \Delta x_{iW}, e_{c,p} = -\Delta x_{iW}$ となり、 $e_c + e_{c,p} = -\Delta x_{iW}$ と $e_p + e_{c,p} = 0$ である。総合誤差 e は0となる。(3) 商品総額の相手国 P_j が Δx_{Tj} 増えたとなると、 $e_c = \Delta x_{Tj}, e_p = 0, e_{c,p} = -\Delta x_{Tj}$ となり、 $e_c + e_{c,p} = 0$ と $e_p + e_{c,p} = -\Delta x_{Tj}$ である。総合誤差 e は0となる。(4) 商品分類 C_i の相手国世界と P_j が同時に Δx_{ij} 増えたとする。すなわち、相手国 P_j については整合性が保たれており商品分類 C_i に誤差が生じている状態である。そのときのそれぞれの誤差は、 $e_c = -\Delta x_{ij}, e_p = 0, e_{c,p} = 0$ 、また、 $e_c + e_{c,p} = -\Delta x_{ij}$ と $e_p + e_{c,p} = 0$ であり、総合誤差 e は $-\Delta x_{ij}$ となる。(5) 相手国 P_j に対して商品総額と C_i が同時に Δx_{ij} 増えたとする。すなわち、商品分類 C_i については整合性が保たれており相手国 P_j について誤差が生じている状態である。そのときの誤差は、 $e_c = 0, e_p = -\Delta x_{ij}, e_{c,p} = 0$ 、また、 $e_c + e_{c,p} = 0$ と $e_p + e_{c,p} = -\Delta x_{ij}$ である。総合誤差 e は $-\Delta x_{ij}$ となる。想定される誤差においてタイプの(2)と(3)は個別分類である商品分類 C_i と相手国 P_j には無関係な誤差であるため総合誤差は0となる特殊な誤差である。タイプ(2)は商品分類 C_i の誤差であるにもかかわらず e_c は0であり、代わりに相手国世界に誤差があるものとして e_p に Δx_{iW} が表示される。 $e_c + e_{c,p}$ と $e_p + e_{c,p}$ はそれぞれ Δx_{iW} と0であり、誤差の状態を正しく表示している。同じようにタイプ(3)は相手国の誤差であるのに e_p は0であり、 $e_p + e_{c,p}$ にその誤差が示されている。個別分類を対象とする一般的な誤差のタイプは(1)、(4)および(5)である。タイプ(1)は商品分類 C_i と相手国 P_j において偶然に生じた異常値や独立に発生する丸め誤差によるものであり、 $e_c + e_{c,p}$ と $e_p + e_{c,p}$ には影響しないため0となることがわかる。タイプ(4)は商品分類のみによる誤差であり、 e_p や $e_{c,p}$ に関わる誤差は0である。タイプ(5)は相手国による誤差であり、 e_c や $e_{c,p}$ に関わる誤差は0である。したがって、商品分類および相手国の誤差を的確に表示するのはそれぞれ $e_c + e_{c,p}$ と $e_p + e_{c,p}$ で

あり、総合誤差 e を加えることで特殊な誤差の状態も識別できる。

(注3) SITC の系列は野田容助「商品分類の改定に伴う対応関係の連結」(古河俊一・野田容助共著『標準国際貿易商品分類と産業分類の対応関係』統計資料シリーズ No.80 アジア経済研究所 1998) の表1を引用、HS の系列については HS-O の部および類は Customs Co-operation Council 作成の *Harmonized Commodity Description and Coding System Explanatory Notes*, first edition (1988) Volume 1、HS-R1 の部および類は World Customs Organization 作成の *Harmonized Commodity Description and Coding System Explanatory Notes*, Second edition (1996) Volume 1 にもとづき筆者作成。HS-O および HS-R1 の項および号は UN から入手した HS-O と SITC-R3 の対応表、HS-R1 と SITC-R3 の対応用のそれぞれにもとづき筆者作成。

(注4) 多くの報告国では1988年以降、各国の商品分類の体系を国連の勧告に従ってHSを基準とした分類体系で編集している。国連の発行する *Commodity Indexes for the Standard International Trade Classification, Revision 3* によればHS-OからSITC-R3への一方方向による変換は可能である。これらの対応関係の中で HS:271000 (Petroleum oils & oils obtained from bituminous minerals, other than crude etc.) は3桁レベルであるSITC-R3:334へしか対応していない。ところが、SITC-R3の一覧表によれば、334の箇所はすべての桁レベルを含む。これらを表示すると、334 (Petroleum oil, preparations)、3341 (Gasolene and other light oils)、33411 (Gasolene including aviation spirit)、33412 (Spirit type

jet fuel)、33419 (Other light petroleum oils)、3342 (Kerosene and other medium oils)、33421 (Kerosene) 33429 (Other medium petroleum oils)、3343 (Gas oils)、3344 (Fuel oils, nes)、3345 (Lubricating petroleum oils) となっている。国連の勧告に従っている報告国では334の個所が3桁の分類になっていることが多い。しかし、HS:271000に含まれる個別例示商品を直接SITC-R3へ対応させた報告国ではこの部分が4桁あるいは5桁の基本項目で編集されている。例外として2桁のみで分類されている国もある。

(注5) データ処理としては対象国11カ国のすべてに対して補正をおこなっている。整合性のある場合には詳細分類における補正は無視される。

【参考文献】

[1] 野田容助『東南アジア貿易統計データ処理(サムチェック作業)について』統計参考資料 74-8 No. 94 アジア経済研究所 1975

[2] ——「商品分類における詳細分類コードの抽出」(野田容助編『世界貿易マトリクスの作成と評価—貿易指数の推計に向けて』調査研究報告書 アジア経済研究所 2002)

[3] ——「貿易からIO24部門分類への変換と世界貿易マトリクス」(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改訂版 アジア経済研究所 2003)