第1部
Part 1

第 1 章

UN Comtrade 貿易データにおける整合性の評価と補正

野田容助

はじめに

UN 統計局（International Trade Statistics Section, UN Statistics Division）編集による UN Comtrade 貿易データは商品分類について階層的に構成されるすべての桁レベル分類コードを含んでいるが、下位桁レベル分類コードの取引額を合計しても必ずしも対応する上位桁レベル分類コードの取引額に一致するとは限らないため、貿易データの整合性を検討する余地があることが知られている。野田・深尾（2005）は UN Comtrade 貿易データにおける整合性の評価および補正についての基本的な考えとして、商品総額と相手国世界を基準として階層的に構成されている貿易データの商品分類の特徴を桁レベル分類コードに基づく取引額という立場から整合性の評価のみならずその整合性の欠如している貿易データの補正方法について紹介している。

本章は野田・深尾（2005）の改訂版であり、貿易データの商品分類における桁レベル分類コードの取引額の整合性を明示的に示すとともに、SITC の基本項目、取引額を考慮した詳細分類である *mdcc* 分類（the most detailed classification code: *mdcc*）、補正された *mdcc* 分類で編集された貿易データのそれぞれ作成する方法を示している。相手国と商品分類の整合性が保証されている貿易データについては *mdcc* 分類コードの取引額の合計は商品総額に一致するが、必ずしも階層的な構造にはなっていないという性質を持っている。UN Comtrade貿易データの桁レベル分類コードおよび *mdcc* 分類コードにもとづく取引額表の評価については、野田・深尾（2005）を含め、野田（2001）、野田（2003によりこれまでたびたび検討してきている。しかし、そこではそれぞれの商品分類コードについての体系を概念的に説明してはいるが正確に定式化してはいない。本章では桁レベル分類コード、*mdcc* 分類コード等についての分類コード体系を定式化して具体例を挙げて示している。商品分類に関しては整合性が保証されていない貿易データの補正について、野田・深尾（2005）は上位の桁レベル分類コードを使用して整合性を高めることを提案しているのに対して、本章では同じ桁レベル分類コードに補正項目を追加して整合性を保証する方法を採用している。

本章は UN Comtrade 貿易データにおける商品分類の特徴、詳細分類としての *mdcc* 分類、相手国が補正されている貿易データに対する商品分類における整合性の評価、商品分類の不整合における補正、から構成されている。

1. 貿易データの商品分類コード

本章で対象とする貿易データは UN の作成による on-line 検索により得られる UN Comtrade 貿易データである。この貿易データは 7 つの分類カテゴリーと 3 つの統計値が存在する。すなわち、分類カテゴリーは報告国（reporter code: *rc*）、年（*y*）、輸出入区分（trade flow code または direction of trade: *d*)、商品分類体系(classification　またはsystem of commodity classification: *sc*)、商品分類コード（commodity code: *c*)、相手国（partner code: *pc*)、数量単位（quantity unit: *qu*）があり、その分類カテゴリーごとに統計値である取引額（value: *v*)、kg を単位とする重量表示の数量（netweight quantity based on kg: *qw*)、数量（supplenmentary quantity: *q*）である。

UN Comtrade 貿易データで使用される貿易商品分類体系は基本的に 2 種類存在し、UN 作成の標準国際商品貿易分類（Standard International Trade Classification: SITC）と関税協力理事会（Customs Co-operation Council）が作成する国際統一商品分類あるいは統一システム（Harmonized Commodity Description and Cording System: HS）である。SITC の商品分類体系はSITC-Original（SITC-O）を初めとして、SITC 改訂版（SITC-Revision）あるいは改訂第1版(SITC-R1)、SITC 改訂第2版(SITC-R2)、SITC 改訂第 3 版（SITC-R3）と最新版の SITC 改訂第 4 版（SITC-R4）がそれぞれ改訂版として存在し、商品総額のもとに 1 桁レベル分類コードから 5 桁レベル分類コードまでの階層的に構成される商品分類コードを含んでいる。SITC 系列の商品分類コードはすべての商品を表わす TOTAL、1 桁レベル分類コードである大分類（Section)、2 桁レベル分類コードの中分類（Division)、3 桁レベル分類コードの小分類（Group)、4 桁レベル分類コードの細分類（Sub-group）があり、さらに詳細分類コードとしての 4 桁レベル分類コードと 5 桁レベル分類コードを含む基本項目（アイテム : Item、または Basic Heading）から構成されている。

4 桁レベル分類コードの中でさらに詳細な内訳が可能なものについてのみ 5 桁レベル分類コードが存在する。したがって、5 桁レベル分類コードは他の分類コードとは構成が異なっており、階層構造になっていないことに注意する必要がある。基本項目において 4 桁レベル分類コードは下位レベルの 5 桁レベル分類コードが存在しないものが対象になっている。それぞれの改訂版により同一分類コードであっても内容が異なるものが存在しているが、分類体系、階層および桁レベルの構成は同一形式である。

HS における商品分類体系は HS1988 年度版の Original（HS-O）を初めとして、HS1996 年度版あるいは改訂第 1 版（HS-R1)、HS2002 年度版あるいは改訂第 2 版（HS-R2）と最新版の HS2007 年度版あるいは改訂第 3 版（HS-R3）となる各改訂版が存在する。HS 系列の商品分類コードは 2 桁レベル分類コードである類（Chapter)、4 桁レベル分類コードの項（Heading)、6 桁レベル分類コードの号（Sub-heading）の他、Chapter を分割して得られる最上位レベルの部（Section）が存在する。

1.1　UN Comtrade 貿易データの取引額表

UN Comtrade 貿易データには商品分類に関して階層構造を構成するそれぞれの桁レベル分類コードの取引額のデータが存在する。k 桁レベル分類コードから構成される商品分類の集まりを、$C(k) = \{c_1(k), \cdots, c_{m(k)}(k)\}$ で表わし、相手国を $P = \{World, p_1, \cdots, p_n\}$ とする。ここで、相手国世界を *World* で表わし、相手国の個数を n、k 桁レベル分類分類コードの個数を $m(k)$ としている。また、$m(0) = 0$ とする。商品分類が SITC のときは、$k = 1, \cdots, 5$ であり、HS のときは $k = 2,4,6$ である。貿易データとして得られる取引額は報告国、年、輸出入区分ごとに、k 桁レベル分類コードで表された商品分類コード $c_i(k)$ と相手国 p_j を、$c_i(k) \in C(k)$ と $p_j \in P$ に対して、

(1-1)　　$$v_{rc,d,y}(c_i(k), p_j)$$

と表わすことができる。報告国、年、輸出入区分、k 桁レベル分類コードを固定して混乱が無いときには簡単に、$v(c_i(k), j)$ あるいは $v(c, j)$ と表すことにする。商品分類について商品総額の *Total* を T で表わし、相手国の世界を *World* を W で表わすことにすれば、$v(T, W)$ は商品総額であり同時に相

手国世界の取引額である。

UN Comtrade 貿易データにおいて商品分類は商品総額とすべての桁レベル分類コードを含んでいて階層構造を形成しているのに対して、相手国については世界合計と個別相手国のみから構成され、地域合計や経済ブロック計のような上位レベルの分類は存在せず階層構造は含まれていないことに注意する必要がある。

UN Comtrade 貿易データを想定したすべての桁レベル分類コードと 3 カ国を相手国とする取引額表が表 1 に示されている。この表は商品分類の SITC-R1 を簡素化して 0 から 0240 までしか存在しないと仮定し、一部を修正して作成された取引額表である。そのため、分類コードは必ずしも SITC-R1 とは一致していない。商品分類は整合性が保証されていない状態を仮定しており、下位レベル分類コードの取引額を合計しても上位レベル分類コードに一致していないものが存在することにも注意すること。この表では特に 4 桁レベル分類コードの 0121 と 0129 には取引額が存在せずに、3 桁レベル分類コードの 012 にのみ取引額が存在すると仮定されている。表 1 において分類コードに影を付けて示されている m が付いて分類コードは第 4 節で説明するように、商品分類の整合性を保証するための補正項目である。簡単化するために相手国については 3 カ国として取引額はすべて同一として、しかも相手国には誤差は存在しないと仮定している。同表における *Error of P* は第 3 節にて説明する。

実際に得られる貿易データは影の部分を取り除いた整合性の欠如した貿易データである。取引額の単位を US$とすれば、商品総額 $v(T,W)$ は 48US$である。一般的には桁レベル分類コードにおける取引額の合計は商品総額には必ずしも一致するとは限らない。表 1 において実際に得られたことを想定した貿易データの k 桁レベル分類コードの取引額合計を *Total*（*k,N*）で表わしている。補正されていない貿易データは N で、*mdcc* 分類コードは *mdcc* と表わされる。同表では 2 桁レベル分類コード以上では一致していないことを示している。補正された貿易データを C で表わし、その k 桁レベル分類コードの取引額合計は *Total*（*k,C*）で表わされている

1.2　階層的な桁レベル分類コード

UN Comtrade 貿易データも含めた一般的な貿易データには商品分類に関して階層構造を構成するそれぞれの桁レベル分類コードが存在する。商品分類コードのすべての集まりを Ω で表すことにする。商品分類が SITC の各改訂版のときは、Ω の要素はすべての商品を表す *TOTAL* と階層構造を持つ 1 桁レベル分類コードから 5 桁レベル分類コードまでの桁レベル分類コードの集まりから構成され、$\Omega = \{Total, C(k)\ k = 1 \cdots 5\}$ となる。*Total* は、$k = 1 \cdots 4$ に対して、$Total = c_1(k) \cup \cdots \cup c_{m(k)}(k)$ である。なお、5 桁レベル分類コードは他の分類コードとは構成が異なっており階層構造になっていない。商品分類コード c が Ω の要素であることを示す識別関数を ξ として、$c \in \Omega$ に対しては $\xi(c) = 1$、$c \notin \Omega$ に対しては $\xi(c) = 0$ とする。さらに、0 から 9 までの 1 桁の数字の集まりを、$A_S = \{0,1,\cdots,9\}$ とする。SITC 各改訂版の k 桁レベル分類コードの集まりは、$k = 1,\cdots,5$ に対して、

$$C(k) = \{\omega \mid \omega \in (C(k-1) \times A_S),\ \xi(\omega) = 1\} \tag{1-2}$$

と表すことができる。ここで、$C(0) = \phi$ としており、記号の $C(k-1) \times A_S$ は $C(k-1)$ と A_S の直積集合を表わしている。$C(k-1)$ に対する下位レベル分類コードの候補としての $C(k-1) \times A_S$ の要素は必ずしもすべてが Ω の属しているとは限らない。そのため、(1-2) 式では ξ を利用して Ω に属するものだけを限定している。

商品分類が HS の Original および各改訂版のときは k 桁レベル分類コードの $C(k)$ を、$k = 2,4,6$ と変更することで SITC 各改訂版と同じように表記

表 1　商品分類 SITC 系列における桁レベル分類コードと相手国の貿易マトリクスの取引額の例

桁　　　　P 1　2　3　4　5	p_1	p_2	p_3	*Error of P*	*World*	基本項目 *N*	*mdcc*分類 *N*	 *C*
Total	16	16	16	0	48			
0 .　　.	16	16	16	0	48			
00　.	4	4	4	0	12			
. 001 .	4	4	4	0	12			
.　　0011	1	1	1	0	3	○	○	○
.　　0012	1	1	1	0	3	○	○	○
.　　0013	1	1	1	0	3	○	○	○
.　　001m	1	1	1	0	3	・	・	●
01　.	9	9	9	0	27			
. 011 .	7	7	7	0	21			
.　　0111	1	1	1	0	3	○	○	○
.　　0112	1	1	1	0	3	○	○	○
.　　0113	1	1	1	0	3	○	○	○
.　　0116	1	1	1	0	3	○	○	○
.　　0118	3	3	3	0	9			
.　　. 01181	1	1	1	0	3	○	○	○
.　　. 01189	1	1	1	0	3	○	○	○
.　　. 0018m	1	1	1	0	3	・	・	●
. 012 .	1	1	1	0	3	・	○	・
.　　0121	0	0	0	0	0	○	・	・
.　　0129	0	0	0	0	0	○	・	・
.　　012m	1	1	1	0	3	・	・	●
. 01m .	1	1	1	0	3	・	・	●
02　.	2	2	2	0	6			
. 022 .	1	1	1	0	3			
.　　0221	1	1	1	0	3	○	○	○
. 023 .	0	0	0	0	0			
.　　0230	0	0	0	0	0	○	・	・
. 024 .	1	1	1	0	3			
.　　0240	1	1	1	0	3	○	○	○
0m　.	1	1	1	0	3	・	・	●
Total（1,*N*）	16	16	16	0	48			
Total（2,*N*）	15	15	15	0	45			
Total（3,*N*）	14	14	14	0	42			
Total（4,*N*）	12	12	12	0	36			
Total（*mdcc,N*）	12	12	12	0	36			
Total（1,*C*）	16	16	16	0	48			
Total（2,*C*）	16	16	16	0	48			
Total（3,*C*）	15	15	15	0	45			
Total（4,*C*）	14	14	14	0	42			
Total（*mdcc,C*）	16	16	16	0	48			

（出所）著者作成。

（注）本表では商品分類の SITC-R1 を基礎として分類コードは *TOTAL* を含め、0 から 0240 までしか存在しないと仮定して作成されている。影の部分は補正項目である。相手国については p_1 から p_3 の 3 カ国でそれらの取引額はすべて同一として、しかも相手国による誤差は存在しないと仮定している。*Total*（*k,N*）は補正されていない貿易データに対する *k* 桁レベル分類コードの取引額の総額、*Total*（*k,C*）は補正済み貿易データのそれを表わす。*mdcc* 分類の *N* と *C* は補正なしおよび補正済みの貿易データを泡和している。基本項目と *mdcc* 分類の○は対象となることを意味する。

できる。したがって、商品分類コードのすべての集まりは、$\Omega = \{Total, C(k) \;\; k = 2,4,6\}$ である。さらに、00 から 99 までの 2 桁の数字の集まりを $A_H = \{00,01,02,\cdots,98,99\}$ とおけば、HS の k 桁レベル分類コードの集まりは、$k = 2,4,6$ に対して、

(1-3) $$C(k) = \{\omega \mid \omega \in (C(k-2) \times A_H), \xi(\omega) = 1\}$$

となる。

具体例として、表 1 の商品分類は SITC に準拠しているので（1-2）式を適用して桁レベ分類コードの作成を試みる。同表から商品分類は補正項目以外を取り出して *Total* も含めたすべての商品分類コードを上から順番に表記すれば、

$$\Omega = \{Total, 0,00,001,0011,\cdots,0240\}$$

となる。1 桁レベル分類コードの集まりは（1-2）式において、$k = 1$、とおいて得られる。$C(0) = \phi$ なので、$C(1)$ は｛0 … 9｝の要素の中から得られる Ω に属するもの集まりであり、

$$C(1) = \{\omega \mid \omega \in (C(0) \times A_S), \xi(\omega) = 1\}$$
$$= \{\omega \mid \omega \in A_S, \xi(\omega) = 1\} = \{0\}$$

となる。$C(1)$ は 1 個の要素から構成される。｛0｝以外の要素は、$\omega \in \{1 \cdots 9\}$ であり、$\omega \notin \Omega$ なので $\xi(\omega) = 0$ となり、$\omega \notin C(1)$ となるからである。次に、2 桁レベル分類コードの集まりは $k = 2$ とおけば、

$$(C(1) \times A_S) = (0 \times A_S) = \{00,01,02,\cdots,09\}$$

なので $C(2)$ は、

$$C(2) = \{\omega \mid \omega \in (0 \times A_S), \xi(\omega) = 1\}$$
$$= \{00, 01, 02\}$$

となる。$C(2)$ は 3 個の要素を含んでいることがわかる。3 桁レベル分類コードは、

$$(C(2) \times A_S) = \{(00 \times A_S) \cup (01 \times A_S)$$
$$\cup (02 \times A_S)\}$$

であり、この中の要素において Ω に属するもの集まりなので、

$$C(3) = \{001, 011, 012, 022, 023, 024\}$$

となり、6 個の要素を含んでいる。同じようにして、4 桁レベル分類コードの集まりを求めることができ、$C(4) = \{0011, \cdots, 0116, 0118, 0121, \cdots, 0240\}$ となる。5 桁レベル分類コードは $C(5) = \{01181, 01189\}$ となる。前者は 13 個、後者は 2 個の要素をそれぞれ含んでいる。

1.3　SITC の基本項目

SITC において詳細分類は基本項目であり、基本項目は 5 桁レベル分類コードが存在しない 4 桁レベル分類コードと 5 桁レベル分類コードの集まりから構成されている。基本分類は SITC の分類体系上の概念であり、分類コードに対応する取引額を考慮していないこと、また、桁レベル分類コードに対する階層的な構造ではないことに注意する必要がある。

k 桁レベル分類コードの $\omega \in C(k)$ において、$m < k$ となる自然数の m に対して、ω の m 桁レベル分類コードを取り出すための関数を $\eta_m(\omega) \in C(m)$ とする。例えば、3 桁レベル分類コードの｛001｝に対しては、$\eta_2(001) = \{00\}$ となる。また、$\eta_m(\phi) = \phi$ である。

5 桁レベル分類コードが存在しない 4 桁レベル分類コードは $C(4) \setminus \eta_4(C(5))$ として得られ、5 桁レベル分類コードは $C(5)$ なので、両者の和集合で表される基本項目は、

(1-4) $$\{C(4) \setminus \eta_4(C(5))\} \cup C(5)$$

となる。ここで、記号の $C(4) \setminus C$ は $C(4)$ から C を取り除くことを意味している。

表 1 の商品分類は SITC-R1 に準拠しているので基本項目が存在するため、（1-4）式を適用して基本項目の作成を試みる。商品分類は補正項目以外を対象とする。$\eta_4(C(5)) = \{0118\}$ なので、基本項目の $C(4) \setminus \eta_4(C(5))$ は $C(4)$ の中から｛0118｝の要素を取り除いたものになる。したがって、基本項目は $C(4)$ における要素の｛0118｝を $C(5)$ の要素の｛01181,01189｝で置き換えたものである。基本項目のすべての用紙は表 1 における基本項目の欄に〇で示され、12 個の 4 桁レベル分類コードと 2

個の5桁レベル分類コードの合計14個の要素から構成されている。

2. 詳細分類としての *mdcc* 分類コード

階層的に構成された商品分類の桁レベル分類コードとは別に、報告国、輸出入区分、年毎の貿易データにおいて取引額が 0 でなく、しかも下位桁レベルの階層に属している分類コードを持たない「詳細分類」コードをアジア経済研究所では *mdcc* 分類コードと呼んでいる。商品分類の詳細分類となる SITC 各改訂版の基本項目が取引額を考慮していないのに対して、*mdcc* 分類コードの特徴は詳細分類コードを選択するときに取引額を考慮していることである。しかも、後述するように世界合計を含む個別相手国を固定すれば整合性が保証されているとき、*mdcc* 分類コードに対応する取引額を合計すると商品総額に一致することである。基本項目では必ずしてその両者が一致するとは限らない。

UN Comtrade 貿易データにおいて SITC-R3 では 3 桁レベル分類コードの 334,673,676 は下位桁レベル分類コードなしに、あるいは下位桁レベル分類コードと共存して使用されていることがある。基本項目を対象としているとこれらの 3 桁レベル分類コードの参照は不可能である。また、野田(2002) において、同じ詳細分類コードとは言っても商品分類が体系的に定義されている基本項目と実際上の分類である *mdcc* 分類コードは必ずしも同一であるとは限らないので貿易データを利用の際にはその両者を混同しないようにとの注意が述べられている。

2.1　*mdcc* 分類コード

商品分類を SITC 各改訂版としたとき、k 桁レベルが c である $k+1$ 桁レベル分類コードの集まりを、$k=1,\cdots,5$ に対して、

(2-1)
$$D_S(c)=\{\omega \mid c\in C(k),\ \omega\in C(k+1),\ \eta_k(\omega)=c\}$$

とする。ここで、$D_S(c)\subset C(k+1)$ であり、当然のことであるが、$c\in C(k)$ に対して $\eta_k(D_S(c))=c$ である。また、(2-1) 式を利用すればすべての $k+1$ 桁レベル分類コードの集まりを表わすことができ、$k=0,\cdots,4$ に対して、

(2-2)
$$C(k+1)=\{\omega \mid \forall c\in C(k),\ \omega\in D_S(c)\}$$

となる[1]。ここで、$\forall c\in C(k)$ と表わされている c は $C(k)$ に含まれているすべての要素であることを意味している。すべての $c\in C(k)$ に対する $D_S(c)$ の集まりがすべての商品分類コードの集まりの Ω なので、

(2-3)
$$\begin{aligned}\Omega\setminus\{Total\}&=C(1)\cup\cdots\cup C(5)\\&=\{\omega \mid \omega\in D_S(c),\forall c\in C(k),\ k=0\cdots4\}\end{aligned}$$

となる。

k 桁レベル分類コードで表された商品分類コードの $c\in C(k)$ と相手国の $p_j\in P$ に対する取引額は (1-1) 式より簡単に $v(c,j)$ と表わすことができる。相手国については、$j\in P\setminus\{World\}$ に対して、$v(c,W)\ge v(c,j)$ なので、取引額を相手国世界の $v(c,W)$ としても一般性を失わない。*mdcc* 分類コードは取引額を考慮したときの詳細分類コードの集まりなので、

(2-4)
$$M_S=\{c \mid c\in C(k),\ k=1\cdots5,\ v(c,W)\neq 0,\ \forall\omega\in D_S(c), v(\omega,W)=0\}$$

となる。もちろん、$v(\phi,W)=0$ としている。ここで重要なことは (2-4) 式からわかるように *mdcc* 分類コードは桁レベル分類コードの k とは無関係に作成されていることである。そのため、*mdcc* 分類コードは必ずしも完全な形での階層構造から構成される桁レベルの分類コードではないことに注意する必要がある。

商品分類を HS の Original および各改訂版としたとき、(2-1) 式に対応させて、k 桁レベルが c である $k+2$ 桁レベル分類コードの集まりを、

$k = 2,4,6$ に対して、

(2-5) $$D_H(c) = \{\omega \mid c \in C(k), \omega \in C(k+2), \eta_k(\omega) = c\}$$

とする。*mdcc* 分類コードは取引額を考慮したときの詳細分類コードの集まりなので、

(2-6) $$M_H = \{c \mid c \in C(k), k = 2,4,6, v(c,W) \neq 0, \forall \omega \in D_H(c), v(\omega,W) = 0\}$$

となる。商品分類が SITC あるいは HS のどちらかにこだわらないときは *mdcc* 分類コードの集まりを簡単に M で表すことにする。

表 1 の商品分類は SITC-R1 を基礎として作成されているので、(2-4) 式を適用して *mdcc* 分類コードの作成を試みる。商品分類は補正項目以外を対象とする。最初は 1 桁レベル分類コードにが *mdcc* 分類コードかどうかを調べるため、$k = 1$ とする。表 3 において $C(1)$ の要素は｛0｝の 1 個が存在するのでこれを $c = \{0\}$ とすれば、$v(c,W) = v(0,W) = 12 \neq 0$ である。1 桁レベルが｛0｝となる 2 桁レベル分類コードの集まりは（2-1）式より、$D_S(0) = \{00,01,02\} \subset C(2)$ となり、3 個の要素が存在する。$C(2)$ の要素の 1 つである｛00｝は $v(00,W) = 3 \neq 0$ なので、$c = \{0\} \notin M$ となる。$c \in C(1)$ において $\omega \in D_S(c)$ が $v(\omega,W) \neq 0$ となるものが 1 つでも存在すれば (2-4) 式を満足しないからである。唯一の $C(1)$ の要素である｛0｝は $\{0\} \notin M$ となるため、1 桁レベル分類コードで M に属するものは存在しない。

同じようにして、2 桁レベル分類コードでは M に属するものは存在せず、3 桁レベル分類コードの中で M に属するものは｛012｝のみであることが示される [2] 。4 桁レベル分類コードについては、｛0118｝以外の 4 桁レベル分類コードは取引額が 0 ではなく、5 桁レベル分類コードが存在しないので M に属し、5 桁レベル分類コードは｛01181, 01189｝は M に属することになる [3] 。

得られた *mdcc* 分類コードのすべての要素は表 1 における *mdcc* 分類コードの補正していない貿易データを表わす N に○で示されている。

2.2 *mdcc* 分類コードの特徴

取引額を考慮した詳細分類としての *mdcc* 分類コードの特徴は世界合計を含む個別相手国を固定すれば整合性が保証されているとき、*mdcc* 分類コードに対応する取引額を合計すると商品総額の $v(T,W)$ に一致することである。商品分類が SITC 各改訂版であるとき、両者が一致することを以下において示す。桁レベル分類コードにおいて 1 桁レベル分類コードの取引額の和は $v(T,W)$ に一致すると仮定する。1 桁レベル分類コードの集まりを、$C(1) = N_1$ とする。また、k 桁レベル分類コードの $c \in C(k)$ に対して、$D_S(c)$ となる $k+1$ 桁レベル分類コードの取引額の和を、

(2-7) $$\sum_{\omega \in D_S(c(k))} v(\omega,W) = v((c(k) \times \bullet),W)$$

とする。最初のステップとして、$k = 1$ とおく。*mdcc* 分類コードの集まりの M は桁レベル分類コードとは無関係なので、M の中で k 桁レベル分類コード集まりを M_k とすれば、

(2-8) $$M_k = \{c(k) \mid c(k) \in N_k, v(c(k),W) \neq 0, v((c(k) \times \bullet),W) = 0\}$$

と表すことができる。$M_k \subset N_k$ であり、M_k の要素の個数を　$m(k)$ とすれば、

$$M_k = \{c_1(k) \cdots c_{m(k)}(k)\}$$

である。N_k は *mdcc* 分類とそうではない分類に分割することができ、$N_k = M_k \cup (N_k \setminus M_k)$ であり、同時に $M_k \cap (N_k \setminus M_k) = \phi$ となる。*mdcc* 分類ではないものの個数を $m'(k)$ とする。N_k の要素の個数は $m(k) + m'(k)$ となり、

$$N_k = \{c_1(k) \cdots c_{m(k)+m'(k)}(k)\}$$

となる。また、整合性が保障されているため、k 桁レベル分類コードのすべての取引額の和は、

(2-9) $$\sum_{c(k) \in N_k} v(c(k),W) = v(T,W)$$

である。

N_k において、$m'(k) > 0$ のとき、*mdcc* 分類コードではない分類コードの集まりは、

(2-10) $$N_k \setminus M_k = \{c_{m(k)+1}(k) \cdots c_{m(k)+m'(k)}(k)\}$$

となり、$N_k \setminus M_k$ の要素を k 桁レベルとする $k+1$ 桁レベル分類コードで取引額の 0 ではないものが存在する。N_k は M_k と $N_k \setminus M_k$ により分割されるため、

(2-11) $$\sum_{c \in N_k} v(c(k), W) = \sum_{c \in M_k} v(c(k), W) + \sum_{c \in N_k \setminus M_k} v(c(k), W)$$

となり、取引額の和を 2 つに分けることができる。N_k のすべての要素が *mdcc* 分類コードである $m'(k) = 0$ のときは、$N_k = M_k$ となる。

$m'(k) > 0$ のときは引き続き、k 桁レベルが $N_k \setminus M_k$ に属する $k+1$ 桁レベル分類コードを N_{k+1} として、

(2-12) $$N_{k+1} = \{\omega \mid c(k) \in (N_k \setminus M_k), \omega \in D_S(c(k))\}$$

とする。その要素の個数を、$m(k+1) +\ \ m'(k+1)$ とする。N_{k+1} は、

$$N_{k+1} = \{c_1(k+1) \cdots c_{m(k+1)+m'(k+1)}(k+1)\}$$

となる。k 桁レベル分類コードで表わされる (2-10) 式は $k+1$ 桁レベル分類コードが存在し、それを表わせば（2-12）式となる。

桁レベル分類コード間に整合性が保証されていれば、k 桁レベル分類コードのすべての $c \in C(k)$ の取引額は、

(2-13) $$v(c(k), W) = v((c(k) \times \bullet), W)$$

である。したがって、k 桁レベル分類コード $c(k)$ の取引額に対するすべての和は

(2-14) $$\sum_{c(k) \in C(k)} v(c(k), W) = \sum_{c(k) \in C(k)} v((c(k) \times \bullet), W) = \sum_{c(k+1) \in C(k+1)} v(c(k+1), W)$$

となり、$k+1$ 桁レベル分類コードのすべての取引額と一致する。さらに、（2-12）式より、

(2-15) $$\sum_{c \in N_k \setminus M_k} v(c, W) = \sum_{c \in N_k \setminus M_k} v((c \times \bullet), W) = \sum_{c \in N_k \setminus M_k} \sum_{\omega \in D_S(c)} v(\omega, W) = \sum_{\omega \in N_{k+1}} v(\omega, W)$$

となる。（2-11）式の右辺の第 2 項目に（2-15）式を代入すれば、

(2-16) $$\sum_{c \in N_k} v(c, W) = \sum_{c \in M_k} v(c, W) + \sum_{c \in N_{k+1}} v(c, W)$$

となる。k を 2 から 4 まで順に置き換えて、この操作を繰り返すことにより、M_1 から M_4 を作成することができる。$k = 5$ のときは、すべての $c \in N_5$ に対して、6 桁レベル分類コードは存在しないので、$m'(6) = 0$ であり、$N_5 = M_5$ となる。これらの一連の操作により M_1 から M_5 が作成される。M は（2-4）式で示される *mdcc* 分類コードの集まりであり、$M = M_1 \cup \cdots \cup M_5$ と表わされる。以上のことから、

(2-17) $$v(T, W) = \sum_{c \in M_1} v(c(1), W) + \sum_{c \in N_2} v(c(2), W) = \sum_{c \in M_1} v(c(1), W) + \cdots = \sum_{k=1}^{5} \sum_{c \in M_k} v(c(k), W)$$

となる。

桁レベル分類コード間に整合性が保証されていることを仮定すれば、*mdcc* 分類コードに対応する取引額を合計すると商品総額の $v(T, W)$ に一致することを確かめられる。商品分類が HS の Original および各改訂版でも同じようにすることで両者の一致を示すことができる。整合性が欠如している貿易データについての *mdcc* 分類コートについては次節にて紹介する。

3. 商品分類における整合性の評価

貿易データの統計値の中で取引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

表2　k 桁レベル分類コードと個別相手国の貿易マトリクスにおける取引額表

P / $C(k)$	p_1		p_j		p_n	*Error of P*	*World*
$c_1(k)$	$v(c_1(k),1)$	$\cdots$	$v(c_1(k),j)$	$\cdots$	$v(c_1(k),n)$	$e_p(c_1(k))$	$v(c_1(k),W)$
:	:				:	:	
$c_i(k)$	$v(c_i(k),1)$	$\cdots$	$v(c_i(k),j)$	$\cdots$	$v(c_i(k),n)$	$e_p(c_i(k))$	$v(c_i(k),W)$
:	:				:	:	
$c_{m(k)}(k)$	$v(c_m(k),1)$	$\cdots$	$v(c_m(k),j)$	$\cdots$	$v(c_m(k),n)$	$e_p(c_m(k))$	$v(c_m(k),W)$
Error of $C(k)$	$e_{c(k)}(1)$		$e_{c(k)}(j)$		$e_{c(k)}(n)$	$e_{c,p}(k)$	b
Total	$v(T,1)$		$v(T,j)$		$v(T,n)$	a	$v(T,W)$

（出所）野田容助「世界貿易マトリクス作成における整合性の評価と補正」（『改訂版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―』SDS No.84 改訂版）の表 1 にもとづき筆者作成。

（注）k 桁レベル分類コードから構成される商品分類の集まりを、$C(k)=\{c_1(k),\cdots,c_{m(k)}(k)\}$ で表わし、相手国の集まりを $P=\{World, p_1,\cdots,p_n\}$ とする。商品総額は *Total* または T、相手国世界は *World* または W である。$c_{m(k)}(k)$ で表わすところ、紙面の都合により、$c_m(k)$ として示されている。相手国による誤差は、$a=e_p(c_\bullet(k))+e_{c,p}(k)$、商品分類による誤差は、$b=e_{c(k)}(\bullet)+e_{c,p}(k)$ である。影の部分は UN Comtrade 貿易データから得られる部分である。

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貿易データの $v(T,W)$ 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。本節は貿易マトリクスについて報告国、輸出入区分、年ごとに商品総額と相手国世界の貿易データを基準値としたときの商品分類および相手国の取引額の整合性を検討している。商品分類については桁レベル分類コード、*mdcc* 分類コードにおける桁レベル分類コード内の整合性を検討すると同時に異なる桁レベル分類コード間の整合性についても検討する。

3.1　桁レベル分類コードの取引額表

桁レベル分類コードの貿易マトリクスの取引額表は同一桁レベル分類コード間における取引表である。k 桁レベル分類コードと個別相手国の貿易マトリクスにおける取引額表が表 2 に示されている。商品分類が SITC 各改訂版のときは、$k=1,2,3,4$ に対する k 桁レベル分類コードと詳細分類としての *mdcc* 分類コードが対象となる。前述したように基本項目では必ずしもそれらの合計は商品総額とは一致するとは限らないため、基本的に商品総額と一致する *mdcc* 分類コードを詳細分類とする。商品分類が HS の Original および各改訂版のときは桁レベルの分類コードは $k=2,4,6$ としたときの k 桁レベル分類コードと詳細分類コードとしての *mdcc* 分類コードが対象となる。

商品総額を基準とすることにより貿易マトリクスの取引額における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、*Error of P*（Partner countries に関する誤差項目）と *Error of* $C(k)$（k 桁レベル分類コードの Commodities に関する誤差項目）が必要となる。UN Comtrade 貿易データをもとにして作成

表 3　k 桁レベル分類コードと個別相手国の貿易マトリクスにおける取引額要約表

P / $C(k)$	P：相手国の合計	*Error of P*	*World*
$C(k)$:商品分類の合計	$v(c_\bullet(k),\bullet)$	$e_p(c_\bullet(k))$	$v(c_\bullet(k),W)$
Error of $C(k)$	$e_{c(k)}(\bullet)$	$e_{c,p}(k)$	$e_{c(k)}(\bullet)+e_{c,p}(k)$
Total	$v(T,\bullet)$	$e_p(c_\bullet(k))+e_{c,p}(k)$	$v(T,W)$

（出所）野田容助（2003）「世界貿易マトリクス作成における整合性の評価と補正」（『改訂版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―』SDS No.84 改訂版）の表 2 に基づき著者作成。
（注）k が 0 のときは商品総額のみしか存在しないため誤差の $e_p(0), e_c(0), e_{c,p}(0)$ はすべて 0 となる。

された貿易マトリクスの取引額表において、商品分類を $c_i(k)$ 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貿易データに関して相手国 j について、$j=1\cdots n$ を合計して、

(3-1)　$$\sum_{j=1}^{n} v(c_i(k),j)=v(c_i(k),\bullet)$$

とする。ここで、$i=1\cdots m(k)$ であり、一般的には必ずしも、$v(c_i(k),W)=v(c_i(k),\bullet)$ であるとは限らない。この原因としては貿易データ固有の誤差、丸め誤差、商品分類や相手国不明の取り扱いによる誤差等が考えられる。商品分類コード $c_i(k)$ の相手国世界の取引額 $v(c_i(k),W)$ を基準としてこの誤差を相手国による誤差 $e_p(k:i)$ とすると、$i=1\cdots m(k)$ に対して、

(3-2)　$$v(c_i(k),W)-v(c_i(k),\bullet)=e_p(c_i(k))$$

と表わされる。表 2 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 $c_i(k)$ に対する取引額の $v(c_i(k),j)$ を $i=1\cdots m(k)$ について合計して、

(3-3)　$$\sum_{i=1}^{n(k)} v(c_i(k),j)=v(c_\bullet(k),j)$$

としても必ずしも相手国 j に対する商品総額に一致するとは限らない。ここで、$j=1\cdots n$ である。この原因も貿易データ固有の誤差や丸め誤差等による誤差等が考えられる。(3-3) 式は表 1 における k 桁レベル分類コードの取引額合計の *Total(k,N)* および *Total(k,C)*に対応する。N は補正されていない貿易データ、C は補正された貿易データである。

商品分類の総額を基準としてこの誤差を k 桁レベル分類コードの商品分類による誤差 $e_{c(k)}(j)$ とすると、$j=1\cdots n$ に対して、

(3-4)　$$v(T,j)-v(c_\bullet(k),j)=e_{c(k)}(j)$$

となる。表 2 において k 桁レベル分類コードにおける商品分類の誤差は *Error of* $C(k)$ の項目に対応する。さらに、同表における *Error of P* の項目と *Error of* $C(k)$ の項目の交点を相手国および k 桁レベル分類コードの商品分類の共通の誤差として $e_{c,p}(k)$ とする。以上のことから、k 桁レベル分類コードにもとづく整合性のある貿易マトリクスは完成する。

次に総合誤差、商品分類による誤差および相手国による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 $i=1\cdots m(k)$ と相手国の $j=1\cdots n$ に対するそれぞれに対する取引額の合計を (3-1) 式および (3-3) 式の表記に対応させて $v(c_\bullet(k),\bullet)$ とする。(3-2) 式において商品分類である $c_i(k)$ の合計を、

(3-5)　$$\begin{aligned} e_p(c_\bullet(k)) &= \sum_{i=1}^{m(k)} e_p(c_i(k)) \\ &= v(c_\bullet(k),W)-v(c_\bullet(k),\bullet) \\ &= e_p(k) \end{aligned}$$

とすると、この誤差は商品分類とは無関係な相手

表4　1桁レベル分類コードと個別相手国の貿易マトリクスにおける取引額表

P $C(k)$	p_1	p_2	p_3	*Error of P*	*World*
0	16	16	16	0	48
Error of　$C(1)$	0	0	0	0	0
Total	16	16	16	0	48

（出所）表1から1桁レベル分類コードの0を取り出し、表2に従って筆者作成。

（注）1桁レベル分類コードでは整合性が保証されているため、*Error of P* と *Error of*　$C(1)$　は共に0である。すなわち、表2における a は0、b は0である。

表5　2桁レベル分類コードと個別相手国の貿易マトリクスにおける取引額表

P $C(k)$	p_1	p_2	p_3	*Error of P*	*World*
00	4	4	4	0	12
01	9	9	9	0	27
02	2	2	2	0	6
Error of　$C(2)$	1	1	1	0	3
Total	16	16	16	0	48

（出所）表1から2桁レベル分類コードの00,01,02を取り出し、表2に従って筆者作成。

（注）相手国の *Error of P* は整合性が保証されているため0であるが、2桁レベル分類コードでは整合性が保証されていないため、*Error of*　$C(2)$　は0ではない。すなわち、表2における a は0、b は3である。

国のみの誤差となる。また、(3-4) 式において相手国である j の合計を、

(3-6)　$e_{c(k)}(\bullet) = v(T,\bullet) - v(c_{\bullet}(k),\bullet) = e_c(k)$

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に共通する誤差は、

(3-7)
$$e_{c,p}(k) = v(T,W) + v(c_{\bullet}(k),\bullet) - v(T,\bullet) - v(c_{\bullet}(k),W)$$

と表わされる。(3-5) 式から (3-7) 式までにより、貿易マトリクスの総合誤差を $e(k)$ とすれば、

(3-8)
$$\begin{aligned} e(k) &= e_c(k) + e_p(k) + e_{c,p}(k) \\ &= v(T,W) - v(c_{\bullet}(k),\bullet) \end{aligned}$$

が得られる。

本章では報告国、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を (3-8) 式、商品分類のみから生じた誤差を、

(3-9)　$e_c(k) + e_{c,p}(k) = v(T,W) - v(c_{\bullet}(k),W)$ と表し、相手国のみから生じた誤差を、

(3-10)　$e_p(k) + e_{c,p}(k) = v(T,W) - v(T,\bullet)$

として表わすことにする。

商品分類の k 桁レベル分類コードの貿易マトリクスの取引額表および要約された取引額表は表2および表3に示され、相手国と商品分類の誤差は $e_p(k), e_c(k), e_{c,p}(k)$ として表示されている。k が0のときは商品総額のみしか存在しないため誤差の $e_p(0), e_c(0), e_{c,p}(0)$ はすべて0となり、総合誤差も $e(0)$ 0 である。k 桁レベル分類コードに基づく貿易マトリクスの表2とその要約表である表3は共に商品総額には k 桁レベル分類コードに対する

表 6　*mdcc* 分類コードと個別相手国の貿易マトリクスにおける取引額表

P / $C(k)$	p_1	p_2	p_3	*Error of P*	*World*
0011	1	1	1	0	3
0012	1	1	1	0	3
0013	1	1	1	0	3
0111	1	1	1	0	3
0112	1	1	1	0	3
0113	1	1	1	0	3
0116	1	1	1	0	3
01181	1	1	1	0	3
01189	1	1	1	0	3
012	1	1	1	0	3
0221	1	1	1	0	3
0240	1	1	1	0	3
Error of $C(mdcc)$	4	4	4	0	12
Total	16	16	16	0	48

（出所）表 1 から *mdcc* 分類コードの *N* を取り出し、表 2 に従って筆者作成。

（注）相手国の *Error of P* は整合性が保証されているため 0 であるが、*mdcc* 分類コードでは整合性が保証されていないため、*Error of* $C(mdcc)$ は 0 ではない。すなわち、表 2 における *a* は 0、*b* は 12 である。

取引額の合計の $v(c_\bullet(k),\bullet)$ ではなく、UN Comtrade 貿易データから得られる商品総額の $v(T,W)$ を採用していることに注意する必要がある。貿易データの整合性の評価において $v(T,W)$ は評価基準として正しいと仮定されており、*k*桁レベル分類コードによる貿易マトリクスの総合誤差は（3-8）式で示されるからである。

表 1 から商品分類の補正項目以外を取り出して桁レベル分類コードによる取引額表を作成する。1 桁レベル分類コードの取引額表は表 4 に示される。この表を表 2 に対応させれば *a* と *b* は共に 0 となり、$v(T,W)=v(c_\bullet(1),\bullet)$ となる。その差も 0 なので、（3-8）式から総合誤差は、$e(1)=48-48=0$ となり、整合性の保証されている取引額表となる。2 桁レベル分類コードの取引額表は表 5 に示される。この表の総合誤差は、$e(2)=48-45=3$ となり、整合性の欠如である取引額表となる。（2-4）式から、$c\in\{00,01,02\}$ に対して、*Error of* $C(2)$の要素は $e_c(2)=48-45=3$ である。省略しているが、同じようにして 3 桁レベルおよび 4 桁レベル分類コードによる取引額表が作成される。*mdcc* 分類コードの取引額表は表 6 に示される。総合誤差は、$e(mdcc)=48-36=12$ となり、整合性の欠如である取引額表となる。$c\in\{0011,0012,\ 0013,\cdots,0240\}$ に対して、$e_c(2)=48-36=12$ である。

3.2　貿易データにおける整合性の評価

貿易データにおいて整合性のある状態ではすべての誤差は存在せず、*k* 桁レベル分類コードに対して、$e_c(k)=0$、$e_p(k)=0$、$e_{c,p}(k)=0$ である。以下、*k* を省略しても一般性を失わないので省略する。それに対して貿易データが整合性に欠けるときの状態は以下の通り 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 に誤差がある例として、整合性のある貿易データに対して $v(c_i,j)$ が $\Delta v(c_i,j)$ 増加した誤差のタイプである。誤差は、$e_c=-\Delta v(c_i,j)$、$e_p=-\Delta v(c_i,j)$、$e_{c,p}=\Delta v(c_i,j)$ となるのに対して、商品分類によ

表 7　商品分類の SITC-R1 を想定した表 1 に基づく桁レベル分類コードの整合性評価表

y	$v(T,W)$	$e(k)$		d_1	d_2	d_3	d_4	d_5	d_6	$e_c(k)+e_{c,p}(k)$		$e_p(k)+e_{c,p}(k)$	
(1 digit level of classification)													
2008	48	0	0.00000	1	0	0	0	0	0	0	0.00000	0	0.00000
(2 digit level of classification)													
2008	48	3	0.06250	0	3	0	0	0	0	3	0.06250	0	0.00000
(3 digit level of classification)													
2008	48	6	0.12500	0	0	6	0	0	0	6	0.12500	0	0.00000
(4 digit level of classification)													
2008	48	12	0.25000	0	0	0	12	0	0	12	0.25000	0	0.00000
(mdcc)													
2008	48	12	0.25000	0	0	1	9	2	0	12	0.25000	0	0.00000

（出所）表 1 から補正項目以外を取り出して著者作成。

（注）y は年、$v(T,W)$ は商品総額、$e(k)$ は総合誤差、$\{d_1 \cdots d_6\}$ は商品分類における k 桁レベル分類コードの個数、$e_c(k)+e_{c,p}(k)$ は商品分類による誤差、$e_p(k)+e_{c,p}(k)$ は相手国による誤差、である。総合誤差、商品分類および相手国の誤差のそれぞれは符号付きの絶対誤差と $v(T,W)$ に対する相対誤差が示される。商品分類を SITC としているため d_6 はすべて 0 となる。表 1 には年度の表記は無いので本表では 2008 年を想定している。

る誤差は $e_c+e_{c,p}=0$ 、相手国による誤差は $e_p+e_{c,p}=0$ となり、これらの誤差は表示されない。総合誤差 e のみが $-\Delta v(c_i, j)$ となる。

⑵ 商品分類 c_i の相手国世界が $\Delta x(c_i,W)$ 増えたときの誤差のタイプは、$e_c=0$ 、$e_p=\Delta v(c_i,W)$ 、$e_{c,p}=-\Delta v(c_i,W)$ となるので、商品分類の誤差は $e_c+e_{c,p}=-\Delta v(c_i,W)$ 、相手国の誤差は $e_p+e_{c,p}=0$ である。総合誤差 e は 0 となる。

⑶ 商品総額に対する相手国 p_j が $\Delta v(T, j)$ 増えたときの誤差のタイプであり、$e_c=\Delta v(T, j)$ 、$e_p=0$ 、$e_{c,p}=-\Delta v(T, j)$ となり、商品分類の誤差は $e_c+e_{c,p}=0$ 、相手国による誤差は $e_p+e_{c,p}=-\Delta v(T, j)$ である。そのときの総合誤差 e は 0 となる。

⑷ 商品分類 c_i の相手国世界と p_j が同時に $\Delta v(c_i, j)$ 増えたときの誤差のタイプである。すなわち、相手国については整合性が保たれており商品分類に誤差が生じている状態である。そのときの誤差は、$e_c=-\Delta v(c_i, j)$ 、$e_p=0$ 、$e_{c,p}=0$ なので、商品分類の誤差は $e_c+e_{c,p}=-\Delta v(c_i, j)$ 、相手国の誤差は $e_p+e_{c,p}=0$ であり、総合誤差 e は $-\Delta v(c_i, j)$ となる。

⑸ 相手国 p_j に対して商品総額と c_i が同時に $\Delta v(c_i, j)$ 増えたときの誤差のタイプである。すなわち、商品分類については整合性が保たれており、相手国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、$e_c=0$ 、$e_p=-\Delta v(c_i, j)$ 、$e_{c,p}=0$ なので、商品分類の誤差は $e_c+e_{c,p}=0$ 、相手国の誤差は $e_p+e_{c,p}=-\Delta v(c_i, j)$ である。総合誤差 e は $-\Delta v(c_i, j)$ となる。

想定される貿易マトリクスの誤差のタイプにおいて、⑵ と ⑶ 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は 0 となる特殊な誤差である。誤差のタイプ ⑵ 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、替わりに相手国世界に誤差があるのものとして e_p に $\Delta v(c_i,W)$ が表示される。商品分類の誤差の $e_c+e_{c,p}$ と相手国の誤差の $e_p+e_{c,p}$ はそれぞれ $\Delta v(c_i,W)$ と 0 であり、誤差の状態を正しく表示している。同じように誤差のタイプ ⑶ 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 ⑴、⑷ および ⑸ である。誤差のタイプ ⑴ 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。誤差のタイプ (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。誤差のタイプ (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。

貿易マトリクス作成における k 桁レベル分類コードの整合性評価表は表 7 に示されている。この表において項目は左から順に、報告国および輸出入区分毎に、年の y、商品総額の $v(T,W)$ 、総合誤差の $e(k)$ 、各 k 桁レベル分類コードの個数は $\{d_1 \cdots d_6\}$ 、商品分類による誤差の $e_c(k)+e_{c,p}(k)$ と相手国による誤差の $e_p(k)+e_{c,p}(k)$ のそれぞれが表示されている。商品分類が SITC 各改訂版のときには 6 桁レベル分類コードは存在しないので d_6 は 0 となる。HS の *Original* および各改訂版のときは d_1, d_3, d_5 は 0 となる。総合誤差、商品分類および相手国の誤差については符号付きの絶対誤差の $e(k)$ ，$e_c(k)+e_{c,p}(k)$ ，$e_p(k)+e_{c,p}(k)$ と $v(T,W)$ に対する相対誤差を共に示している。

表 1 から商品分類の補正項目以外を取り出して桁レベル分類コードによる整合性の評価表を作成する。表 2 において商品分類による誤差は a、相手国による誤差は b で表わされている。1 桁レベル分類コードについては表 4 から a と b は共に 0 であり、前節から総合誤差は 0 であることが知られているので表 7 のように整合性の保証された評価表となる。分類コードの個数は {0} の 1 個なので、1 桁レベル分類コードを表わす d_1 の列に 1 が記載されている。2 桁レベル分類コードについては表 5 から a と b はそれぞれ 0 と 3 である。相手国には誤差が生じていないので総合誤差はすべて商品分類によるものである。誤差は 3 であることが知られているので表 7 のように整合性に欠如があるという評価表となる。総合誤差には絶対誤差の 3 と相対誤差の 0.0625 が示されている。分類コードの個数は {00,01,02} の 3 個なので、2 桁レベル分類コードを表わす d_2 の列に 3 が示されている。3 桁レベルおよび 4 桁レベル分類コードについても同じようにして作成される。

mdcc 分類コードは分類コードの個数の表示が桁レベル分類コードのそれとは異なる。*mdcc* 分類コードの個数は 3 桁レベル分類コードの {012} の 1 個、4 桁レベル分類コードの {0011} から {0240} までの 9 個、5 桁レベル分類コードの {01181, 01189} の 2 個がそれぞれ d_3, d_4, d_5 の列に示される。

3.3　異なる桁レベル分類コード間の整合性

桁レベル分類コードと個別相手国の貿易マトリクスの取引表およびその要約表である表 2 および表 3 で示されているように商品総額と相手国世界の取引額は k 桁レベル分類コードにおける商品総額の $v(c_\bullet(k),\bullet)$ ではなく、$v(T,W)$ が利用されている。同一 k 桁レベル分類コード内で整合性が保証されていないときには、その取引額総額と商品総額との差の $v(T,W)-v(c_\bullet(k),\bullet)$ が誤差となる。実はこの差は整合性の評価における総合誤差であり、(3-8) 式で示されている。それと同時に階層的に構成されている商品分類では異なる桁レベル分類コード間においても誤差が生じる。そのため、同一桁レベル分類コードについての整合性評価と同時に、異なる桁レベル分類コードについて整合性評価も必要となる。

k 桁レベル分類コードの $c(k) \in C(k)$ に対する取引額の $v(c(k),W)$ と下位レベルとなる $k+1$ 桁レベル分類コードの集まりである $D_S(c(k))$ の取引額の和 $v((c(k)\times\bullet),W)$ の差を両者の符号付絶対誤差として、

(3-11)　　$v(c(k),W)-v((c(k)\times\bullet),W)=\alpha(c(k))$

とする。また、c に対する相対誤差を、

表 8　k 桁レベル分類コードの取引額とその下位レベルのそれの和の整合性

桁レベル	c	p	q	$\alpha(c)$
$k+1$	(空白)	$v(c_\bullet(k),W)$	$v(c_\bullet(k+1),W)$	$\alpha(c_\bullet(k))$
$k+1$	*	$\sum_{c\in C(k)^*} v(c,W)$	$\sum_{c\in C(k+1)^*} v(c,W)$	$\sum_{c\in C(k)^*} \alpha(c)$
$k+1$ ・	$c_1(k)$	$v(c_1(k),W)$	$v((c_1(k)\times\bullet),W)$	$\alpha(c_1(k))$
$k+1$	$c_1(k),0$	.	$v((c_1(k),0),W)$	.
：	：	：	：	：
$k+1$	$c_1(k),9$	.	$v((c_1(k),9),W)$	.
：	：	：	：	：
$k+1$ ・	$c_{m(k)}(k)$	$v(c_{m(k)}(k),W)$	$v((c_{m(k)}(k)\times\bullet),W)$	$\alpha(c_{m(k)}(k))$
$k+1$	$c_{m(k)}(k),0$	.	$v((c_{m(k)}(k),0),W)$	.
：	：	：	：	：
$k+1$	$c_{m(k)}(k),9$	.	$v((c_{m(k)}(k),9),W)$	.

（出所）著者作成。

（注）k と c は k 桁レベル分類コードの $c\in C(k)$ を表わす。c が空白のときはすべての商品、*のときは誤差を含むすべての商品を表わす。p、q、$\alpha(c)$ の表記については本章の 3.2 を参照すること。k 桁レベル分類コードの評価を意味するため、誤差は「k・」で示され、この後ろに「k」として $D_S(c_i)$ の取引額の内訳が q に示されるが、p と $\alpha(c)$ は存在しないので「.」とする。

$$\alpha(c(k))/v(c(k),W)=\beta(c(k)) \quad \text{(3-12)}$$

とする。$k+1$ 桁レベル分類コードのすべての取引額の和である $v((c_\bullet(k)\times\bullet),W)$ は（3-11）式と（2-2）式により、

$$\begin{aligned} v((c_\bullet(k)\times\bullet),W) &= \sum_{c\in C(k)} \sum_{\omega\in D_S(c)} v(\omega,W) \\ &= v(c_\bullet(k+1),W) \end{aligned} \quad \text{(3-13)}$$

となる。（3-11）式に対する商品総額における絶対誤差は、

$$\begin{aligned} & v(c_\bullet(k),W)-v(c_\bullet(k+1)),W) \\ &= \alpha(c_\bullet(k)) \end{aligned} \quad \text{(3-14)}$$

と表わされる。この式が本節で紹介する桁レベル分類コード間の誤差である。

桁レベル分類コードの取引額とその下位レベルのそれの和の整合性は表 8 に示されている。同表において取引額は相手国世界を対象として、商品分類が SITC の各改訂版のときは分類コードの桁レベルの k は 1 から 5 まで、HS の Original および各改訂版は k は 2,4,6 となる。商品分類を SITC 改訂版としたとき、同表における桁レベル分類コードの取引額の p と q の表記は以下のように 3 通りに分かれる。HS についても基本的には同一である。

（1）表 8 において c が空白のとき、k 桁レベル分類コードの商品総額を基準としたとき、$k+1$ 桁レベル分類コードの商品総額との比較により整合性を評価する。同表の p は k 桁レベル分類コードにおける商品総額であり、c_i について $i=1\cdots n(k)$ の取引額の合計となる商品総額の $v(c_\bullet(k),W)$ である。q は $k+1$ 桁レベル分類コードの商品総額の $v((c_\bullet(k)\times\bullet),W)$ であるが、これは（3-13）式から $v(c_\bullet(k+1),W)$ となるので、絶対誤差の $p-q$ は（3-14）式で示される。対象となる k は 0 から 3 までである。k が 0 のときに（3-14）式において、$c_\bullet(0)$ となるが、これを商品総額の *Total* あるいは T で表わすことにする。そのときの取引額は、$v(c_\bullet(0),W)=v(T,W)$ となる [5]。表 8 では $\alpha(c_\bullet(k))$ が 0 でないときに $k+1$ 桁レベル分類コードによる誤差の内訳が表示される。k が 4 のときの表示は省略する。というのは、4 桁レベル分類コードの

$c \in C(4)$ について、$D_S(c) = \phi$ となるものが存在するため、4 桁レベル分類コードの取引額のすべての和より 5 桁レベル分類コードの取引額の和が小さくなり絶対誤差が 0 とならない。そのため、すべての 5 桁レベル分類コードが誤差として表示されるからである。

(2) 表 8 において c が*のときは、k 桁レベル分類コードにおける、$\alpha(c(k)) \neq 0$ についての絶対誤差の和を表わし、

(3-15)　$$C(k)^* = \{c \mid c \in C(k), v((c \times \bullet), W) \neq 0, \alpha(c) \neq 0\}$$

としたときの k 桁レベル分類コードの取引額の合計額を、

$$p = \sum_{c \in C(k)^*} v(c, W)$$

として、$k+1$ 桁レベル分類コードのそれを、

$$q = \sum_{c \in C(k)^*} v(c \times \bullet), W)$$

とする。絶対誤差の $\alpha(c)$ は $p-q$ で示される。対象とする k は 0 から 4 までで、$\alpha(c) \neq 0$ についてのみ表示されている。

(3) 表 8 において c が $c_i(k) \in C(k)$ のとき、p は k 桁レベル分類コード $c_i(k)$ の取引額 $v(c_i(k), W)$ である。q は $D_S(c_i(k))$ の取引額の和を表わし、$v((c_i(k) \times \bullet), W)$ である。絶対誤差の $\alpha(c_i(k))$ は (3-11) 式で示される。対象となる k は 0 から 4 までであり、$\alpha(c_i(k)) \neq 0$ についてのみ表示されている。(3) についての表示方法は 2 種類の項目が存在し、最初は分類コード $c_i(k)$ の評価を意味する内容が「k･」で示され、この後に「k」として $D_S(c_i(k))$ の取引額の内訳のすべてが q として示される。内訳については p と $\alpha(c)$ は存在しないので欠損値を表わす「.」が表記される。

表 8 において取引額は相手国世界を利用しているため、(3-11) 式の絶対誤差 $\alpha(c_\bullet(k))$ と (3-9) 式の整合性における商品分類の誤差の関係は、

(3-16)　$$\{e_c(k+1) + e_{c,p}(k+1)\} - \{e_c(k) + e_{c,p}(k)\} = \alpha(c_\bullet(k))$$

と表わされる [6]。ここで、k は 0 から 3 までを対象とする。相手国を補正した表 1 を利用して両者の誤差を比較したのが表 9 である。総合誤差の $e(k)$ は表 7 から得られ、絶対誤差の $\alpha(c_\bullet(k))$ は表 1 から、k 桁レベル分類コードのすべての商と相手国世界の取引額は $v(Total(k, N), W)$ であり、それが $v(c_\bullet(k), W)$ に一致することからその差で得られる。表 9 から (3-16) 式が成り立つのを確認できる。

3.4　異なる桁レベルによる整合性の例

表 1 から商品分類の補正項目を対象外として取り出して桁レベル分類コード桁レベル分類とその下位レベル分類コードの取引額の和の整合性評価表を作成したのが表 10 である。表 8 において k を 0 とすれば、整合性の評価基準となっている商品総額の $v(T, W)$ と 1 桁レベル分類コードの取引額の和の比較となる。商品分類コード c の空白はすべての商品を表わし、p の $v(c_\bullet(0), W)$ は $v(T, W)$ であり、q は $v(c_\bullet(1), W)$ である。表 1 から p は 48、q の $v(c_\bullet(k), W)$ は $v(Total(k, N), W)$ に対応するため、$v(Total(1, N), W)$ は 48 となり、両者の差となる絶対誤差は $\alpha(c_\bullet(0)) = p - q = 0$ となる。表 10 の 1 行目にこの結果が示されている。1 桁レベル分類コードは整合性が保証されていることを意味し、これ以上の表記はおこなわれない。

表 8 において k が 1 のときは 1 桁レベル分類コードの取引額と 2 桁レベル分類コードの取引額の和の比較である。商品分類コード c が空白のとき、p は $v(c_\bullet(1), W)$、q は $v(c_\bullet(2), W)$ である。表 1 から p は 48、q の $v(c_\bullet(2), W)$ は $v(Total(2, N), W)$ に対応するため 45、絶対誤差は $\alpha(c_\bullet(1)) = p - q = 3$ となる。この誤差は 1 桁レベル分類コードにおける $\{0\}$ の取引額からその下位レベル分類コードの $D_S(0) =$ $\{00, 01, 02\}$ の取引額の和を差し引いたものである。表 10 の 2 行目にこの結果が示されている。表 8 において、絶対誤差のすべての和を表示するのが、c の*である。この例では 1 桁レベル分類コードは 0 しか存在していないので c の空白と*

表 9　桁レベル分類コード間の誤差と整合性評価における商品分類に関わる誤差の関係

$k+1$	$v(c_\bullet(k),W)$	$v(c_\bullet(k+1),W)$	$\alpha(c_\bullet(k))$	$e_c(k)+e_{c,p}(k)$	$e_c(k+1)+e_{c,p}(k+1)$	d
1	48	48	0	0	0	0
2	48	45	3	0	3	3
3	45	42	3	3	6	3
4	42	36	6	6	12	6

（出所）桁レベル商品総額の $v(c_\bullet(k),W)$ は表 1 の $v(Total(k,N),W)$ から、整合性評価の商品分類による誤差の $e(k)$ は表 7 からそれぞれ引用して著者作成。

（注）k は桁レベル分類コード、$\alpha(c_\bullet(k))$ は（3-14）式から求められる。$e_c(k)+e_{c,p}(k)$ は商品分類の関わる整合性の誤差、d = $\{e_c(k+1)+e_{c,p}(k+1)\}$-$\{e_c(k)+e_{c,p}(k)\}$である。

表 10　表 1 における桁レベル分類コードとその下位レベル分類コードの取引額の和の整合性

k+1	c	p	q	α(c)	k+1	c	p	q	α(c)
1		48	48	0	4	0012	.	3	.
2		48	45	3	4	0013	.	3	.
2	*	48	45	3	4・	011	21	21	0
2・	0	48	45	3	4	0111	.	3	.
2	00	.	12	.	4	0112	.	3	.
2	01	.	27	.	4	0113	.	3	.
2	02	.	6	.	4	0116	.	3	.
3		45	42	3	4	0118	.	9	.
3	*	27	24	3	4・	012	3	0	3
3・	00	12	12	0	4	0121	.	0	.
3	001	.	12	.	4	0129	.	0	.
3・	01	27	24	3	4・	022	3	3	0
3	011	.	21	.	4	0221	.	3	.
3	012	.	3	.	4・	023	0	0	0
3・	02	6	6	0	4	0230	.	0	.
3	022	.	3	.	4・	024	3	3	0
3	023	.	0	.	4	0240	.	3	.
3	024	.	3	.	5	*	9	6	3
4		42	36	6	5・	0118	9	6	3
4	*	15	9	6	5	01181	.	3	.
4・	001	12	9	3	5	01189	.	3	.
4	0011	.	3	.					

（出所）著者作成。

（注）表の構成は表 8 にしたがって作成されている。影の部分は本来ならば表示されないことになっている $\alpha(c)$ が 0 となっている分類コード c であるが、表を見やすくするためにすべての分類コードを表示している。k が 5 のときは c の空白部分は省略され、「*」以降が表示される。

は同じ誤差となる。誤差の内訳は「1・」において示されるが、誤差が 1 個であるため「1　*」と「1・」の誤差は一致している。さらに、引き続いて、誤差生じている 2 桁レベル分類コードとその取引額が示される。表 10 にその結果が示されている。

表 8 において k が 2,のときは、2 桁レベル分類コードの取引額と 3 桁レベル分類コードの取引額の和の比較である。2 桁レベル分類コードの 3 個

表 11　桁レベル分類コードにおける *mdcc* 分類コードの誤差

K	$\sum_{c\in N_k} v(c(k),W)$	$\sum_{c\in M_k} v(c(k),W)$	$m(k)$	$\sum_{c\in N_{k+1}} v(c(k+1),W)$	$\sum_{c\in N_k \backslash M_k} \alpha(c)$
0	48	0	0	48	0
1	48	0	0	45	3
2	45	0	0	42	3
3	42	3	1	36	3
4	36	27	9	6	3
5	6	6	2	0	0
合計	.	36	12	.	12

（出所）著者作成。

（注）表頭の項目は（3-18）式を参照すること。左辺と右辺がそれぞれ分かれて表示されている。α は桁レベル分類コードにおける *mdcc* 分類の誤差である。

の{00,01,02}に対応する 3 桁レベル分類コードはそれぞれ $D_S(00)$ ={001}、$D_S(01)$ ={001,012}、$D_S(02)$ ={022,023,024}から構成され、6 個存在する。表 8 において c が空白のとき、p は $v(c_\bullet(2),W)$、q は $v(c_\bullet(3),W)$ である。表 1 から p と q はそれぞれ、$v(Total(2,N),W)$ の 42、$v(Total(3,N),W)$ の 36、に対応し、絶対誤差は $\alpha(c_\bullet(2)) = p - q = 6$ となる。誤差が生じているのは 2 桁レベル分類コードの｛01｝の 1 個であり、その取引額と下位レベル分類コードの $D_S(01)$ の取引額の和の間で生じている。誤差の 6 は誤差の内訳は「3・」において示されるが、誤差が 1 個であるため、「2　（空白）」、「2　*」、「2・」の誤差は一致している。

表10に結果が示されている。この表では整合性の保証されている{00,02}についての絶対誤差は0であるが、桁レベル分類コード間の関係を見やすくするように影を付けて表わしている。kが3,4についても同様である。表7で示された桁レベル分類コードの整合性評価における総合誤差と表10で示された桁レベル分類コードの絶対誤差は必ずしも一致していないが、表9においてその違いを示している。

3.5　不整合な貿易データの*mdcc*分類

貿易データが *mdcc* 分類コードで分類されおり、しかも整合性が保証されていないとき、k 桁レベル分類コードの $c_i \in C(k)$ は（3-11）式において、$\alpha(c_i) \neq 0$ となる分類コードが少なくとも 1 個は存在する。整合性が保証されているときは $\alpha(c_i)$ は 0 となるので、桁レベル分類コード間の取引額の比較の一般的な式は（3-11）式で表される。（3-11）式において c をすべての商品の *Total* とすれば、その取引額の $v(T,W)$ と 1 桁レベル分類コードの集まり $C(1)$ に対する取引額の和の差額が一致しないときには、（3-14）式より、$N_1 = C(1)$ なので、

$$v(T,W) - \sum_{c\in N_1} v(c(1),W) = \alpha(c_\bullet(0)) \tag{3-17}$$

となる。（3-11）式を（2-11）式の右辺の第 2 項に代入すれば、（2-15）式より、$k = 1 \cdots 4$ に対して、

$$\sum_{c\in N_k} v(c(k),W) = \sum_{c\in M_k} v(c(k),W) + \sum_{c\in N_{k+1}} v(c(k+1),W) + \sum_{c\in N_k \backslash M_k} \alpha(c(k)) \tag{3-18}$$

となる。（3-18）式において $k = 1$ とおき、これを（3-17）式に代入して、

$$\begin{aligned} v(T,W) &= \sum_{c\in N_1} v(c(1),W) + \alpha(c_\bullet(0)) \\ &= \sum_{c\in M_1} v(c(1),W) + \sum_{c\in N_2} v(c(2),W) \\ &+ \sum_{c\in N_1 \backslash M_1} \alpha(c(1)) + \alpha(c_\bullet(0)) \end{aligned}$$

と表すことができる。この操作を k が 2 から 4 になるまで繰り返せば、

(3-19)
$$v(T,W)=\sum_{k=1}^{5}\sum_{c\in M_k}v(c(k),W)+\alpha(c_\bullet(0))$$
$$+\sum_{k=1}^{4}\sum_{c\in N_k\setminus M_k}\alpha(c(k))$$

が得られる。

整合性の保障されていない貿易データでは（3-19）式により商品合計は桁レベル分類コードの *mdcc* 分類コードと誤差の和で表されることが確認できる。*mdcc* 分類における桁レベル分類コードにおける絶対誤差の合計は、

(3-20)
$$\alpha(c_\bullet(0))+\sum_{k=1}^{4}\sum_{c\in N_k\setminus M_k}\alpha(c(k))$$

として示される。1 桁レベル分類コードの誤差は $\alpha(c_\bullet(1)=v(T,W)-v(c_\bullet(1),W)$ であり、$k=1\cdots 4$ に対する k 桁レベル分類コードの誤差は（3-18）式の右辺の第 3 項目で示される。整合性が保障されている貿易データでは（3-20）式で表されている誤差が 0 となることであり、この誤差を 0 とおけば整合性の保証された（2-17）式に一致する。したがって、*mdcc* 分類コードにおける誤差評価の一般形は（3-19）式で表される。

表 1 の例として *mdcc* 分類における桁レベル分類コードの誤差の具体例を示す。（3-17）式において、$k=1$ とする。$N_1=C(1)=\{0\}$ なので、

(3-20)
$$\sum_{c\in N_1}v(c(1),W)=v(0,W)=48$$

である。（3-17）と（3-20）式から 1 桁レベル分類コードの誤差は、$\alpha(c_\bullet(1))=$48-48=0、となる。1 桁レベル分類コードにおいて *mdcc* 分類コードは存在しないので、$M_1=\phi$ となり、

(3-21)
$$\sum_{c\in M_1}v(c(1),W)=0$$

であり、$N_1=M_1\cup(N_1\setminus M_1)$ から、$N_1\setminus M_1=\{0\}$ であるので、2 桁レベル分類コードの N_2 は、

$$N_2=\{\omega\,|\,c\in(N_1\setminus M_1),\,\omega\in D_S(c)\}$$
$$=\{00,01,02\}$$

として得られ、その取引額は、

(3-22)
$$\sum_{c\in N_2}v(c(2),W)=v(00,W)+v(01,W)$$
$$+v(02,W)=45$$

である。（3-18）式において $k=1$ とおけば、1 桁レベル分類コードの誤差は、

$$\sum_{c\in N_1\setminus M_1}\alpha(c(1))=\text{48-0-45=3}$$

となる。2 桁レベル分類コードの誤差を求めるため、（3-18）式において、$k=2$ とする。$M_2=\phi$、$N_2\setminus M_2=\{00,01,02\}$ であるので、$N_3=\{001,011,012,\cdots,024\}$ となる。N_3 の取引額は、

(3-23)
$$\sum_{c\in N_3}v(c(3),W)=v(001,W)+\cdots$$
$$+v(024,W)=42$$

である。同じようにして 2 桁レベル分類コードの誤差は、

$$\sum_{c\in N_2\setminus M_2}\alpha(c(2))=\text{45-0-42=3}$$

となる。3 桁レベル分類コードについては、（3-18）式において、$k=3$ とする。$M_3=\{012\}$、$N_3\setminus M_3=\{001,\ 011,\ 022,\ 023,\ 024\}$ であるので、$N_4=\{0011,0012,\ \cdots\ ,024\}$ となり、その誤差は 3 となる。同じようにして、4 桁レベル分類コードは、$M_4=\{0011,\ \cdots\ ,0016,0121,0129,\ \cdots\ ,0240\}$、$N_4\setminus M_4=\{0018\}$ であるので、$M_5=N_5=\{01181,01189\}$ となり、その誤差は 3 となる。

表 11 に桁レベル分類コードにおける *mdcc* 分類の誤差が示されている。この表において $m(k)$ は M_k における要素の個数、α は（3-17）式と（3-18）式で表されている *mdcc* 分類コードの桁レベルの誤差である。この表から 3 桁レベル分類コードの *mdcc* 分類コードの個数は 1、4 桁レベル分類コードは 9、5 桁レベル分類コードは 2 であり、それぞれの取引額の和は 3、27、6 であることが示される。また、桁レベルによる誤差は 1 桁レベルから 4 桁レベルまですべて 3 となっており、その合計は 12 である。したがって、表 1 で表される貿易データを *mdcc* 分類コードに編集し直すと商品総額の 48 うち *mdcc* 分類で表される部分が 36、誤差が 12 と

なる。表 7 で示される整合性の評価表との関係では、(3-10) 式の右辺の第 2 項目の $v(c_\bullet(k),W)$ を、

$$\sum_{k=1}^{5}\sum_{c\in M_k} v(c(k),W)=v(c_\bullet(mdcc))$$

と置き換えることにより、商品分類による誤差は、

(3-24) $$e_c(mdcc)+e_{c,p}(mdcc)=v(T,W)-v(c_\bullet(mdcc),W)$$

と表すことができる。したがって、表 7 の *mdcc* の行の商品分類に関わる誤差の $e_c(k)+e_{c,p}(k)$ と表 11 の α の合計が一致する。桁レベル分類コードから構成されていない *mdcc* 分類コードの個数は表7と表において3,4,5桁レベル分類コードのそれぞれにおいて一致している。

4. 貿易データにおける不整合の補正

UN Comtrade 貿易データにおける不整合の状態は総合誤差である (3-8) 式の $e(k)$ が、$k=1\cdots4$ および *mdcc* 分類コードに対して 0 ではないことであり、補正は総合誤差が 0 となるように補正項目を追加して整合性を保証することである。総合誤差は商品分類と相手国による誤差から構成されているため、補正は両者に対する補正が必要となる。補正は表 5 において誤差の項目である *Error of P* の列である (3-2) 式の $e_p(c_i(k))$、*Error of* $C(k)$ の行である (3-4) 式の $e_{c(k)}(j)$、その交点になる (3-7) 式の $e_{c,p}(k)$ をそれぞれ推計して補正項目として追加することで可能となる。特殊な不整合の状態でない限り相手国についての誤差は UN Comtrade 貿易データで使用されている相手国の分類不明を表す国コードの 899 へ相手国の誤差である (3-2) 式を対応させる。すなわち、$k=1\cdots5$ に対して、商品分類の $c_i(k)$, $i=1\cdots m(k)$ の相手国の誤差を、

(4-1) $$v(c_i(k),899)=e_p(c_i(k))$$

として補正項目を追加して補正する。

この補正により相手国の誤差は 0 となるので相手国の補正は完了する。以下、簡単化するために相手国は既に補正されている貿易データに対して商品分類を補正する方法を示す。一般的には相手国と商品分類が補正された後にその交点である相手国と商品分類の共通の誤差が最後に補正される。商品分類の補正は相手国が補正済みの貿易データを対象とする。数量単位はすべての桁レベル分類コードにおいて同一であると仮定する。

桁レベル分類コードから構成される階層構造の商品分類体系において整合性のある状態というのは下位レベル分類コードの取引額合計が対応する上位レベル分類コードの取引額と一致していることであり、商品総額を基準とすれば各桁レベル分類コードの取引額合計が商品総額と一致する。商品分類の整合性に欠ける状態は下位桁レベル分類コードの取引額合計が対応する上位レベル分類コードの取引額と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。野田・深尾（2005）は階層構造を持つ桁レベル分類コードにおいて、k 桁レベル分類コードの取引額と k 桁レベル分類コードが同一であるすべての k+1 桁レベル分類コードの取引額合計の差が大きいときには k 桁レベルの分類コードを使用して整合性を高めている。この補正方法は複数個の $k+1$ 桁レベル分類コードを削除して 1 個の k 桁レベル分類コードを追加しているため、補正後に分類コードの個数は減少する。それに対して本章の補正の方法は k 桁レベル分類コードの取引額と k 桁レベル分類コードが同一であるすべての k+1 桁レベル分類コードの取引額合計の差が大きいときには k+1 桁レベルの分類コードに補正項目を追加して整合性を保つことを提案している。

4.1　商品分類の追加による補正

商品分類がSITC各改訂版のときの桁レベル分類コードは、$k=1\cdots5$ に対して $C(k)$ で表わされる。k桁レベル分類コードの $c\in C(k)$ と相手国世界に対する取引額は $v(c,W)$ である。商品分類コードの k桁レベルが c である $k+1$ 桁レベル分類コードの

集まりは（2-1）式で示されている $D_S(c)$ であり、k桁レベルがcである $k+1$ 桁レベル分類コードの取引額の和は（3-11）式である。$k+1$ 桁レベル分類コードによる補正は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。符号を付けた絶対誤差は（3-11）式である。絶対誤差に対する補正基準を α^* とすれば、商品分類コードcに対する絶対誤差に対する補正は、

(4-2) $$\alpha(c) \geq \alpha^*$$

であるとき、$k+1$ 桁レベル分類コードにおける補正項目の分類コードを、

(4-3) $$(c \times m) \in C(k+1)$$

とし、同時にその取引額を、

(4-4) $$v((c \times m), W) = \alpha(c)$$

とし、補正項目を（2-1）式に追加することでおこなう。すなわち、

(4-5) $$D_S(c)^* = D_S(c) \cup (c \times m)$$

とする。この式により、k 桁レベルが c である $k+1$ 桁レベル分類コードの取引額の和を、

$$\sum_{\omega \in D_S(c)^*} v(\omega, W) = v((c \times \bullet\bullet), W)$$

とする。（4-2）式を満たさないときは補正はおこなわない。微小な取引額もさることながら、特に $\alpha(c) < 0$ については補正項目の取引額が負となるため、有効な取引額として利用できないからである。補正をしないときの絶対誤差は（3-11）式の $\alpha(c)$ であるのに対して、補正したときのの絶対誤差は、

$$\alpha(c)^\bullet = x(c, W) - x((c \times \bullet\bullet), W) = 0$$

となる。この補正基準α^*を絶対補正係数といい、$0 \leq \alpha^*$ である。商品分類コードcに対する相対誤差は（3-12）式である。cに対する絶対誤差に対する補正は相対誤差の補正基準を β^* とすれば、

(4-6) $$\beta(c) \geq \beta^*$$

となるとき、絶対誤差の補正と同じように（4-3）式の補正項として $D_S(c)$ に $k+1$ 桁レベル分類コードを追加することでおこなわれる。（4-6）式を満たさないときには補正はおこなわない。β^* を相対補正係数といい、$0 \leq \beta^* \leq 1$ である。

補正係数の α^* が0のときは β^* も0となり、(4-4) 式において $\alpha(c)$ が0ではないすべてのk+1桁レベル分類コードを補正項目として追加することになる。一方では、$\beta^* = 1$ のときは丸めの誤差により必ずしも正しくはないが、一般に貿易データでは、$v(c, W) - v((c \times \bullet), W) > 0$ であるので（3-12）式において $v(c, W) < 0$ となる。このようなことは起こり得ないのでk+1桁レベル分類コードの追加はおこなわれない。このことは補正をしないことを意味する。このようにして α^* および β^* の値を適当に選択することによりk+1桁レベル分類コードに対する追加をするときと同じように追加しないときにも利用できる。

すべての $c \in C(k)$ に対する補正項目が追加された $D_S(c)^*$ の集まりが整合性のある商品分類コードのすべての集まりとなる。(2-3）式に対応させてこれを Ω^* とすれば、

(4-7) $$\Omega^* \setminus \{Total\} = \{\omega \mid \omega \in D_S(c)^*, \forall c \in C(k) \quad k = 1 \cdots 4\}$$

である。補正基準の値にもよるが、（3-8）式にもとづく貿易データの整合性評価表では総合誤差が補正前に比較して小さくなっている。

商品分類体系がHSでも同じように補正項目の追加による補正が可能である。HSの商品分類において桁レベル分類コードは、$k = 0,2,4,6$ に対して $C(k)$ で表わされる。SITC系列の（4-3）式に対応させればHS系列ではk桁レベルがcである $k+2$ 桁レベル分類コードの集まりは (2-5) 式の $D_H(c)$ であり、その要素におけるそれぞれの取引額の和を、

$$\sum_{\omega \in D_H(c)} v(\omega, W) = v((c \times \bullet), W)$$

とする。HS系列の4桁レベル分類コードで表される商品分類コードを $c \in C(4)$ とすれば、4桁レベル分類コードがcである6桁レベル分類コードの取引額の和は $v((c \times \bullet), W)$ 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（3-11）式に対応させるとその絶対誤差は $\alpha(c)$ で示され、その補正

表 12　補正された表 3 に基づく桁レベル分類とその下位レベルの和の整合性

k	c	p	b	$\alpha(c)$	k	c	p	q	$\alpha(c)$
1		48	48	0	4	＊	15	15	0
2		48	48	0	4・	001	12	12	0
2	＊	48	48	0	4	0011	.	3	.
2・	0	48	48	0	4	0012	.	3	.
2	00	.	12	.	4	0013	.	3	.
2	01	.	27	.	4	001m	.	3	.
2	02	.	6	.	4・	012	3	3	0
2	0m	.	3	.	4	0121	.	0	.
3		45	45	0	4	0129	.	0	.
3	＊	27	27	0	4	012m	.	3	.
3・	01	27	27	0	5	＊	9	9	0
3	011	.	21	.	5・	0118	9	9	0
3	012	.	3	.	5	01181	.	3	.
3	01m	.	3	.	5	01189	.	3	.
4		42	42	0	5	0118m	.	3	.

（出所）著者作成。
（注）表 10 に同じ。

基準を α^* とすれば、（4-2）式を満たすとき、補正はこの不等式を満たす6桁レベルを分類コードとして追加することである。この不等式を満たさないときは追加をおこなわない。また、相対誤差による補正基準は相対誤差を（3-13）式として、その補正基準を β^* とするとき、（3-13）式を満たす4桁レベルの分類コードに対して6桁レベル商品分類コードを補正子項目として追加する。HS系列の2桁レベル分類コードについても同じように取り扱うことが可能である。HSについても（4-7）式と同じように、すべての $c \in C(k)$ に対する補正された $D_H(c)^*$ の集まりが整合性のある商品分類コードのすべての集まりとなり、

$$\Omega^* \setminus \{Total\} = \{\omega \mid \omega \in D_H(c)^*, \forall c \in C(k), k = 0,2,4\} \quad (4\text{-}8)$$

である。

4.2　整合性の欠如に対する補正

整合性の欠如している貿易データの補正に対して本節では絶対誤差の補正基準を α^* を 0 と設定する。このとき、β^* も 0 となり、(4-2) 式において $\alpha(c)$ が 0 ではないすべての k+1 桁レベル分類コードを補正項目として追加することになる。整合性評価表の表 9 から 1 桁レベル分類コードは総合誤差が 0 であり、整合性が保証されていることがわかる。整合性のある桁レベル分類コードの補正はおこなわれない。2 桁レベル分類コードから 4 桁レベル分類コードについては、整合性の欠如を意味する総合誤差および商品分類に関わる誤差は 0 ではないことが示される。相手国に関わる誤差が 0 であるため総合誤差はすべて商品分類によるものと判断できる。

整合性の欠如している貿易データの補正は表 11 に基づいて補正項目を作成することでおこなう。2 桁レベル分類コードについては表 10 の 2 桁レベル分類コードで示されている誤差の 3 は表 11 において 1 桁レベル分類コードと 2 桁レベル分類コードの間で生じた誤差として示され、前者の{0}の取引額と後者の $D_S(0)$ の要素となる {00,01,02} の取引額の和との間に 3 の差が生じていることを示している。絶対補正基準は α^* は 0 なので、(3-1)

表 13　補正された表 1 に基づく桁レベル分類コードの整合性評価表

y	$v(T,W)$	$e(k)$		d_1	d_2	d_3	d_4	d_5	d_6	$e_c(k)+e_{c,p}(k)$		$e_p(k)+e_{c,p}(k)$	
(1 digit level of classification)													
2008	48	0	0.00000	1	0	0	0	0	0	0	0.00000	0	0.00000
(2 digit level of classification)													
2008	48	0	0.00000	0	4	0	0	0	0	0	0.00000	0	0.00000
(3 digit level of classification)													
2008	48	3	0.06250	0	0	7	0	0	0	3	0.06250	0	0.00000
(4 digit level of classification)													
2008	48	6	0.12500	0	0	0	14	0	0	6	0.12500	0	0.00000
(mdcc)													
2008	48	0	0.00000	0	1	2	9	3	0	0	0.00000	0	0.00000

（出所）表 1 から補正項目を含めすべてを取り出して著者作成。

（注）表 7 に同じ。

式から、$v(0,W)-v((0\times\bullet),W)=\alpha(0)$、となり、$\alpha(0)>\alpha^{*}=0$ なので（4-1）式を満足する。このときには（4-2）式に従って補正項目の $(c\times m)=\{0m\}$ が作成され、その取引額として $v(0m,W)=\alpha(0)=3$ となる。補正は $D_S(0)$ にこの補正項目を追加することでおこなわれ、（4-3）から、$D_S(0)^{*}=D_S(0)\cup\{0m\}=$ {00,01,02,0m} となる。$v((0\times\bullet\bullet),W)=48$ となり、商品総額に一致する。補正された結果は表 11 の k が 2 の箇所に示されているとおり $\alpha(0)^{\bullet}=0$ である。表 12 に補正された貿易データにおける整合性の評価が示されており、2 桁レベル分類コードの総合誤差が 0 になっているのを確かめることができる。分類コードの個数の d_2 は 4 であり、補正前の貿易データの整合性評価表の表 9 における d_2 の 3 から補正項目を追加した 1 個が増えている。

表 10 によれば、3 桁レベル分類コードは $D_S(00)$, $D_S(01)$, $D_S(02)$ から構成され、その中で絶対誤差が 0 でないものは $D_S(01)$ であり、$\alpha(01)=3$ は（4-1）式を満足する。このときには（4-2）式に従って補正項目の {01m} が作成され、その取引額は $v(01m,W)=\alpha(01)$ =3 となる。（4-3）式から、

$$D_S(01)^{*}=D_S(01)\cup\{01m\}=\{011,012,01m\}$$

となる。4 桁レベル分類コードおよび 5 桁レベル分類コードについても同じようにして補正をおこなうことができ、前者は {001m,012m}、後者は {0118m} が補正項目となる。分類コードの個数は補正前後の整合性評価表の表 7 と表 12 を比較することにより、d_3 は 1 個、d_4 は 2 個それぞれ増加している。

補正された貿易データのすべての分類コードは表 1 に示されている。同表おいて影を付けているのが補正項目であり、補正項目を含めないすべての分類コードが Ω、補正項目を含めたそれが Ω^{*} である。補正前後の貿易データに対して *mdcc* 分類コードの一覧は表 3 の *mdcc* 分類の欄に示され、前者は N、後者は C の列にそれぞれ○で表わされている。補正前後で異なる分類コードは {012} と補正項目に含まれる分類コードである。補正前の貿易データにおいて {012} は $D_S(012)=$ {0121,0129} と下位桁レベル分類コードを持っているにもかかわらずその取引額の和となる $v((012\times\bullet),W)$ が 0 であるため、（2-4）式を満足し *mdcc* 分類コードに属している。しかし、補正された貿易データでは $D_S(012)^{*}=D_S(012)\cup\{012m\}$ で、$v(012m,W)$ =3 なので、$D_S(012)^{*}$ の取引額の和の $v((012\times\bullet),W)$ は 0 ではない。そのため {012} は *mdcc* 分類コードに属さない。その代わりに 4 桁レベル分類コードの {012m} が *mdcc* 分類コー

ドとなる[6]。補正前後の貿易データに対する *mdcc* 分類コードの整合性の評価が表 7 と表 12 にそれぞれ示されている。前者の総合誤差は 12 であるのに対して、後者の総合誤差は 0 であり、(2-4) 式で示されているように *mdcc* 分類コードの取引額の和は商品総額に一致している。*mdcc* 分類コードの個数は補正前後の整合性評価表の表 9 と表 14 を比較することにより、d_2 は 1 個、d_4 は 2 個、d_5 は 1 個それぞれ増加している。d_3 については {012} がなくなり、{012m} が追加されたため個数の増減は変化していない。

表 12 において補正された貿易データであるにも関わらず 3 桁レベルおよび 4 桁レベル分類コードには依然として総合誤差に 3 と 6 が残ったままである。補正は桁レベル分類コード間においておこなわれているため、3 桁レベル分類コードの取引額合計が 2 桁レベル取引額合計に一致するように 3 桁レベル分類コードの補正項目が作成される。表 9 において k が 3 の箇所に示されているように、2 桁レベル分類コードの取引合計は $v(c_{\bullet}(2),W)$ の 45 であり、3 桁レベルのそれは 42 である。補正項目は 3 桁レベル分類コードの取引額合計に 3 を追加することで作成される。商品総額は 48 であるため、補正後にも 3 の誤差が残っている。4 桁レベル分類コードは 3 桁レベル分類コードの取引総額が基準となるため商品総額の 48 から 3 桁レベル分類コードの取引額の和の 42 を引いた 6 が誤差として残っている。補正された貿易データの利用に *mdcc* 分類コードを用いるのは、桁レベル分類コードではその取引額の和が商品総額には必ずしも一致しないのに対して、(2-14) 式で示されているように *mdcc* 分類コードの取引額の和は必ず商品総額に一致するからである。

おわりに

本章は野田・深尾（2005）の改訂版であり、UN 作成によるUN Comtrade貿易データの中で、特に商品分類に関する整合性評価と可能な限り実施できる補正の方法を説明することを目的としている。さらに、桁レベル分類コード、*mdcc*分類コード、整合性の評価および補正に関する概念を定式化してこれらの関連をまとめて表現できるように組み替えている。貿易データで使用される商品分類は階層的な構造を持つ桁レベル分類コードから構成されており、それぞれの桁レベル分類コードの集まりはSITCの各改訂版は（1-2）式で示され、HSのOriginalおよび各改訂版は（1-3）式で示される。階層構造が成り立たないSITCの基本項目については（1-4）式で示される。

階層的に構成された商品分類コードとは別に、報告国、輸出入区分、年毎の貿易データにおいて取引額が0でなく、しかも下位桁レベルの階層に属している分類コードを持たない「詳細分類」コードが*mdcc*分類コードある。*mdcc*分類コードはSITCの各改訂版は（2-4）式で表わされ、HSのOriginalおよび各改訂版は（2-6）式で示される。商品分類がSITC各改訂版のときには*k*桁レベル分類コードに含まれる*mdcc*分類コードは（2-8）式の M_k で示され、（2-12）式より、$k=1\cdots4$ に対して、$\eta_k(N_{k+1}\cap M_k)=\phi$ となる。このことは*k*桁レベル分類コードを $c\in M_k$ とすれば、$k+1$ 桁レベル分類コードで*k*桁レベルがcとなるものが存在しないことを保証している。HSのOriginalおよび各改訂版についても同様である。相手国と商品分類の整合性が保証されている貿易データについては*mdcc*分類コードの取引額の合計は商品総額に一致することは（2-15）式により示される。*mdcc*分類コードは必ずしも階層的な構造になっていないことに注意する必要がある。

貿易データにの整合性については報告国、年、輸出入区分を固定したときの階層的な桁レベル商品分類と個別相手国をもとにした貿易マトリクスから生ずる誤差は、商品分類に関わる誤差は (3-9) 式、相手国に関わる誤差は（3-10）式、商品分類と相手国を含む総合誤差は（3-8）式で表わされ、これらの誤差の状態をまとめたのが整合性の評価表である。k 桁レベル分類コードの整合性評価表

は表 7 および表 12 で示されている。

相手国の整合性が保証されている貿易データについて数量単位が同一であるときには商品分類の整合性を維持させるための補正は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。符号を付けた絶対誤差は（3-11）式、相対誤差は（3-12）式である。絶対誤差に対する補正基準を$\alpha*$とすれば、商品分類コードcに対する絶対誤差に対する補正は（4-2）式、相対誤差は（4-6）式を満足するとき、$k+1$桁レベル分類コードにおける補正項目の分類コードを$(c\times m)\in C(k+1)$、同時にその取引額を$\alpha(c)$として補正項目を追加することでおこない、（4-5）式となる。補正基準を満たさないときには補正はおこなわれない。一般的には相手国の整合性は保証されていないので、貿易データの補正は相手国の補正と商品分類の補正に付け加えて、補正のための絶対誤差α^*を0としないときは四捨五入による丸めの誤差を含むため、最終的な相手国と商品分類に対する共通の補正により処理される。

貿易データの桁レベル分類コードの整合性と補正については商品分類コードの他に相手国および数量単位や数量に関しても必要であり現在検討中の課題である。補正については絶対誤差と相対誤差による補正基準を設定しているがこれをどの程度にすればいいかという課題は依然として残されている。

---

[1] （2-2）式が成り立つことを示す。まず、$D_S(\phi)=C(1)$である。というのは、（2-2）式において$k=0$とすれば、$C(0)=\phi$なので、$c=\phi$であり、$\omega\in C(1)$に対して$\eta_0(\omega)=\phi$となり、（2-2）式を満足するからである。次に、$0<k<5$については、$C(k)=\{c_1\cdots c_{n(k)}\}$とする。$i=1\cdots n(k)$に対して、$c_i\in C(k)$とすれば、$D_S(c_i)$はk 桁レベルが$c_i$となる$k+1$桁レベル分類コードの集まりである。したがって、

$$D_S(c_i)=\{\omega\mid\omega\in(c_i\times A_S),\xi(\omega)=1\}$$

となる。$D_S(c_1)$から$D_S(c_{n(k)})$を集めたものが$k+1$桁レベル分類コードのすべての集まりである。すなわち、$C(k+1)$は$\forall c\in C(k)$に対する$D_S(c)$の集まりであるのでその和をとれば、

$$\begin{aligned}\bigcup_{i=1}^{n(k)} D_S(c_i) &= \{\omega\mid j\in A_S,\omega\in((c_1\times j)\cup\cdots\\ &\cup(c_{n(k)}\times j)),\xi(\omega)=1\}\\ &=\{\omega\mid j\in A_S,\omega\in(C(k)\times j),\xi(\omega)=1\}\\ &=C(k+1)\end{aligned}$$

となり、

$$C(k+1)=\{\omega\mid\forall c\in C(k),\omega\in D_S(c)\}$$

である。このことから（2-2）式が示される。

[2] 2 桁レベル分類コードが *mdcc* 分類コードかどうかを調べるため、$k=2$とする。表 1 より$C(2)$の要素は｛00,01,02｝である。最初は｛00｝が *mdcc* 分類コードかどうかを調べる。$v(00,W)=3\neq 0$であり、（2-1）式より、$D_S(00)=\{001\}\subset C(3)$となる。$C(3)$の要素である｛001｝は$v(001,W)=3\neq 0$なので（2-6）式を満足せず、$\{00\}\notin M$となる。2 番目は｛01｝を調べる。$v(01,W)=7\neq 0$であり、$D_S(01)=\{011,012\}\subset C(3)$となる。$C(3)$の要素である｛011｝は$v(011,W)=6\neq 0$なので、$\{01\}\notin M$となる。3 番目の｛02｝は$v(02,W)=2\neq 0$であり、$D_S(02)=\{022,023,024\}\subset C(3)$である。$C(3)$の要素の｛022｝は$v(022,W)=1\neq 0$なので、$\{02\}\notin M$となる。$C(2)$のすべての要素が（2-2）式を満足しないので、2 桁レベル分類コードでMに属するものは存在しない。3 桁レベル分類コードが *mdcc* 分類コードかどうかを調べるため、$k=3$とする。2 桁レベル分類コードと同じようにして$\{001,011,022\}$はMの要素ではないことがわかる。｛012｝は$v(012,W)=1\neq 0$であり、$D_S(012)=\{0121,0129\}\subset C(4)$である。この 2 つの要素は$v(0121,W)=0$、$v(0129,W)=0$ある。さらに、$D_S(012)$の 2 つの要素は共に下位レベル分類コードが存在しないため、$D_S(0121)\ =D_S(0129\}=\phi$であり、$v(\phi,W)=0$である。したがって、｛012｝は（2-4）式を満足することから、Mの要素であることが示される。｛023｝は$v(023,W)=0$であると同時に、$D_S(023)=\{0230\}$において、$v(0230,W)=0$となる。（1-5）式を満足しないので｛023｝$\notin M$となる。以上のことから、3 桁レベル分類コードの中でMに属するものは｛012｝のみである。

[3] 4 桁レベル分類コードについては$C(4)$の要素の｛0011｝は、$v(0011,W)=1\neq 0$であり、下位レベルと

なる 5 桁レベル分類コードが存在しないので無条件に M に属する。それは、$D_S(0011)=\phi$ であり、$v(\phi,W)=0$ となることから（2-4）式を満足するからである。同じようにして、{0118} 以外の 4 桁レベル分類コードは取引額が 0 ではなく、5 桁レベル分類コードが存在しないので M に属する。{0118} は $v(0118,W)=2\neq 0$ であり、$D_S(0118)=$ {01181, 01189} $\subset C(5)$ となる。$C(5)$ の要素の 1 つが $v(01181,W)=1\neq 0$ となるため（1-7）式を満足しないため、$\{0118\}\notin M$ である。5 桁レベル分類コードは $c\in C(5)$ に対して $v(c,W)\neq 0$ であり、しかも下位レベル分類コードが存在しないため、$v(01181,W)=1\neq 0$ と $v(01181,W)=1\neq 0$ であることから {01181,01189} は M に属することになる。

4　k 桁レベル分類コードの集まりの $C(k)$ において、$C(0)=\phi$ であるが、$c_\bullet(0)$ は商品分類コードとしての商品総額の*Total*あるいは*T*であるので間違いのないように注意する必要がある。$v(C(0),W)=v(\phi,W)=0$ であるのに対して、$v(c_\bullet(0),W)=v(T,W)$ である。

5　（3-14）式が成り立つことを示す。（3-9）式において $k+1$ から k を引けば（3-14）式が導かれる。（3-9）式において $k=1$ とおけば、

$$e_c(1)+e_{c,p}(1)=v(T,W)-v(c_\bullet(1),W)=\alpha(c_\bullet(0))$$

となり（3-14）式を満足する。これは $e_c(0)$ と $e_{c,p}(0)$ が共に0であることを利用している。

6　表1において分類コード c の取引額の $v(c,W)$ が0であるにもかかわらず取引額表に計上されているのは数量が0でない分類コードであり、{0121,0129,023,0230} がそれに相当する。実際の貿易データにおいて*mdcc*分類コードは取引額だけではなく数量も考慮して定義されている。*mdcc*分類コードを表現する（2-4）式は数量を考慮すれば、SITC各改訂版では、

$$M_S=\{c\mid c\in C(k), k=1\cdots 5, v(c,W)\neq 0,$$
$$\forall\omega\in D_S(c),$$
$$v(\omega,W)=0, qw(\omega,W)=0, q(\omega,W)=0\}$$

となる。ここで、*qw*と*q*はそれぞれUN Comtrade貿易データにおけるkgを数量単位とする数量、数量単位に基づく数量を表している。本章では簡単化するために数量を考慮していないため、補正前の貿易データにt対する*mdcc*分類コードは012としているが、数量を考慮すれば012は*mdcc*分類コードではない。補正後の*mdcc*分類コードは012mとなり、これは012と同一内容である。敢えて補正項目として追加する必要もないが、補正の追加を説明するためにこのままにしている。

参考文献

野田容助（2001）「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『商品分類の変換に伴う貿易統計の変換』統計資料シリーズ（SDS）No.83 アジア経済研究所）

———（2002）「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（開発研究部 2001-Ⅲ-12）アジア経済研究所）

———（2003）「世界貿易マトリクス作成における整合性の評価と補正」（野田容助編『改訂版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（SDS）No.84 Revised アジア経済研究所）

野田容助編（2005）『東アジア諸国・地域の貿易指数―作成から応用までの基礎的課題―』 統計資料シリーズ（SDS）No.88 アジア経済研究所

野田容助・黒子正人（2006）「貿易指数の作成と応用に向けた諸課題」（野田容助・黒子正人編『長期時系列における貿易データと貿易指数の作成と応用』調査研究報告書（開発研究センター 2005-Ⅱ-04）アジア経済研究所）

野田容助・深尾京司（2005）「貿易マトリクス作成における整合性の評価―新および旧 AID-XT 基礎データにもとづいて―」（野田容助編『東アジア諸国・地域の貿易指数―作成から応用までの基礎的課題―』 統計資料シリーズ（SDS）No.88 アジア経済研究所）