Chapter3 Evaluation and Revision to Consistency of Compiling World Trade Matrices

Chapter 3

Evaluation and Revision to Consistency of Compiling World Trade Matrices

NODA Yosuke

The national and regional economies that make up the world economy are incorporated into its dynamism on the basis of international interdependence, and their mutual close ties have been growing stronger. For purposes of international comparison, the world trade statistics generated by the OECD (referred to as the OECD World Trade Statistics) and those generated by the United Nations (the UN World Trade Statistics) are usually used. When focusing specifically on the value of the commodities traded as part of world trade statistics, system of commodity classification includes not only respective commodity classification codes but also commodity total, and system of partner country includes not only respective partner country but also world. For this reason, handling the value with commodity total and partner country world as normative values lets us investigate the consistency of the various commodity classification and partner countries to find out whether the sum values of the individual commodity classification codes or the individual partner country data are consistent with the normative values. The value arrived at by subtracting the total value from the normative value is referred to as "error," and the method that evaluates the error by taking the sum of the individual commodity classification codes or the individual partner countries. This method which looks at the consistency of the commodity classification codes by comparing the existing commodity total value and the sums of the values is called the "sum check."

This chapter looks at the examination of these world trade matrices in terms of sum checks of the commodity classification and partner countries for each reporting country and import/export classification. I will also describe a method for adjusting the data for reporting countries for which consistency cannot be guaranteed so that there is consistency in the sense required for a sum check.

The commodity classification used in trade statistics is composed of trade classification codes divided into strata at the digit level based on the commodity total. The Institute of Developing Economies refers to the hierarchically composed classification codes that have no lower level classification code as "most detailed classification codes" (*mdcc*). In this chapter, we treat the sum total value of the transactions for the *mdcc*s in the commodity classifications as an evaluation criterion consistent conformity with the commodity total values.

If $C = \{C_1 \cdots C_m\}$ is the commodity classification composed of *mdcc*s, and

$P = \{P_1 \cdots P_n\}$ is the partner countries composed of the individual partner countries, then the data for the total value of the transactions in the trade statistics for each reporting country (*rc*), year (*y*), and type of trade (*d*) can be represented as:

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

If the reporting country, year, and type of trade are fixed, this value is represented simply as $x_{i,j}$. As mentioned previously, since the commodity total value (*Total*) for each commodity classification is included in the actual trade statistical data, the cpommodity classification for purposes of revision is $C^* = \{Total, C\}$, and since the world is included in the partner countries in terms of the sum of partner countries, it is generally the case that $P^* = \{World, P\}$.

The trade matrix that was generated by correlating the partner countries to columns and the commodity classifications to rows is shown in Table 1. The parts that were obtained as actual data are shaded and displayed as x_{ij}. With respect to the commodity classifications *i*, if the commodity total value is represented by *T*, and the world with respect to the partner countries *j* is represented by *W*, then x_{TW} is both the commodity total value and at the same time, the value of the transactions with the world. Having this x_{TW} as the criterion makes it possible to evaluate the consistency of the trade matrices. In order to ensure consistency in the trade matrices, we need the error items for each of the partner countries and the commodity classifications, *the error of P* (partner countries) and *the error of C* (commodities), but for those data we represent them as *e*.

The transaction values in the trade matrices generated on the basis of statistical trade data will not necessarily conform to those for the partner countries worldwide, even if we fix the commodity classifications as C_j and add up the partner countries. In general, if we assume that the sum of the partner countries *j* with respect to the data obtained is

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

then it is $x_{iW} \neq x_{i\bullet}$. The cause is believed to be rounding errors and to transactions for which the identity of the partner country is unclear. If the criterion is data on trade with the world x_{iW} for a commodity classification C_i, and this error is assumed to be error by partner country $e_p(i)$, then it is represented as

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

In the world trade matrix in Table 1, it corresponds to the *error of P* with respect to the partner countries. We can also say the same about the commodity classifications. If the partner country is set as P_j, then the sum of the commodity classifications will not necessarily be the same as the total value of the commodities. If we assume that

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

then $x_{Tj} \neq x_{\bullet j}$. If we take the total value of the commodity classifications as the norm and assume that this error is the error $e_c(j)$ due to the commodity classifications, then

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

This corresponds to the *error of C* items in the trade matrix in Table 1. Furthermore, the intersection of the *error of P* items and the *error of C* items in Fig. 1 is common error $e_{c,p}$ of the partner countries and the commodity classifications. Based on the above, we can complete a consistent world trade matrix on the basis of the most detailed classification codes.

Next, let the sum of partner countries and the commodity classifications in the data obtained in

order to find the estimated value of the error be

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

If the sum of the commodity classifications in eqn. (2) is

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

then this error is the error due to the partner countries without any connection to the commodity classifications. Moreover, if the sum total for the partner countries in eqn. (3) is

$$(5) \qquad e_c(\bullet) = \sum_{i=1}^{m} (x_{Tj} - x_{\bullet j}) = x_{T\bullet} - x_{\bullet\bullet}$$

this error is the error due to the commodity classifications without any connection to the partner countries. The common error of the partner countries and the commodity classifications is represented as

$$(6) \qquad e_{c,p} = (x_{TW} + x_{\bullet\bullet}) - (x_{T\bullet} + x_{\bullet W})$$

In this chapter, the error of partner countries is defined,

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

as the sum of eqn (4) and eqn (6). Also the error of commodity c;assification is defined,

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

as the sum of eqn (5) and eqn (6).

If, based on eqn. (4) to eqn. (6), we let the total error for the trade matrix be *e*, then we find that

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

From this, the trade matrix error e based on the commodity classifications when the reporting country, the year, and type of trade are fixed is represented as the sum of the error e_c arising from the commodity classifications alone, the error e_p arising from the partner countries alone, and the error $e_{c,p}$ arising from both the commodity classification and the partner countries. It is therefore possible to represent a consistent table as

$$rc, y, d, e_p[mdcc], e_c[mdcc] + e_{c,p}[mdcc], \\ e_p[mdcc] + e_{c,p}[mdcc]$$

in a sum check based on the *mdccs*. If it is determined that there is a problem of consistency only with the commodity classifications, the following analysis considers the error due to the sum at the commodity classification digit level. $\{d_1 \cdots d_6\}$ based on eqn. (1), we fix the reporting country, type of trede, years, and the world, and use x_{iW} with respect to $i \in C$.

In the SITC commodity classification system, the sum for the digit level is represented as $x_{iW}[k]$ with respect to $k = 0, \cdots, 4$. Of course, the total value of the commodity is $x_{iW}[0] = x_{TW}$. Since the criterion for the sum check is the commodity total value, the absolute error of evaluation a_k and relative error of evaluation b_k for the respect digit levels is for $k = 1 \cdots 4$,

$$(8) \qquad a_k = \mid x_{TW} - x[k] \mid \\ b_k = \mid x_{TW} - x[k] \mid / x_{TW}$$

the sum with respect to the *mdccs* is $x_{\bullet W}$, so the error of evaluation ratio is

$$(9) \qquad a_D = \mid x_{TW} - x_{\bullet W} \mid \\ b_D = \mid x_{TW} - x_{\bullet W} \mid / x_{TW}$$

When it is determined that there is a problem with consistency in the sum at the digit level, we need to conduct a further analysis of error in the commodity classification at the digit level. When the commodity classifications are according to the SITC classification system, we use a different representation than in eqn. (1) and represent the value of the transactions as

$$(10) \qquad x(rc, d, y, W : i_1, i_2, i_3, i_4, i_5)$$

For $k = 1 \cdots 5$, $i_k = 0 \cdots 9$. Eliminating eqn. (10), we assume $x(i_1, i_2, i_3, i_4, i_5)$. Moreover, eqn. (10) represents the value of the transactions for the k-digit

level commodity classification code $\{i_1,\cdots,i_k\}$ as $x(i_1,\cdots,i_k)$. The total value of the k-digit level classification codes $\{i_1,\cdots,i_k\}$ obtained by summing up the k+1 level classification codes is represented as

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

which makes clear the difference between them Despite the fact that 5 digits level classification codes exist in the United Nations trade statistics compiled according to the SITC classification system, some of the summed up transaction values do not conform to the transaction values according to the 4 digits level classification codes. It is known through evaluation of consistency at various digit levels of commodity classifications that the higher the number of digits, the greater the consistency. Therefore, when the difference between the transaction values obtained when 5 digits level classification codes are summed at the 4 digits level and the total value of transactions in the 4 digits level classification codes is large, we are faced with the need to increase consistency by using the 4 digits level classification code, not the 5 digits level code. The process of increasing consistency by this method is called 5 digits level classification code correction.

Correction criterion is determined by number of 4 digits level classification codes and the two types of error, absolute error and relative error, that occur with respect to the data of the 5 digits level classification codes compiled to the 4 digits level. If the correction based on the absolute error is a, and we use eqn. (11), the result is

$$a=|x(i_1,i_2,i_3,i_4,i_5)-x(i_1,i_2,i_3,i_4,\bullet)|$$

the correction criterion is

(11) $$a\geq a^*$$

Even if there exists a 5 digits level commodity classification code, the 4 digits level classification codes $\{i_1,\cdots,i_4\}$ that satisfy eqn. (11), their 4 digits level is used as the classification code. If they do not satisfy eqn. (11), we use the 5 digits level commodity codes. Here a^* is referred to as the absolute correction coefficient, and it is $a^*\geq 0$.

Moreover, if we let the correction norm based on the relative error b is

$$b=\frac{|x(i_1,i_2,i_3,i_4,i_5)-x(i_1,i_2,i_3,i_4,\bullet)|}{x(i_1,i_2,i_3,i_4,\bullet)}$$

then the correction norm is

(12) $$b\geq b^*$$

Even if a 5 digits level commodity classification code exists, 4 digits level commodity classification codes $\{i_1,\cdots,i_4\}$ that satisfy eqn. (12) are used as 4 digits level classification codes. Here b^* is referred to as the relative correction coefficient, and it is $0\leq b^*\leq 1$.

If the correction coefficient is a^*=0, then b^* is also 0, and this means replacing all the 5 digits level classification codes with 4 digits level codes. On the other hand, when b^*=1, there is no substitution of 5 digits level codes with 4 digits level codes. By appropriately selecting the values of a^* and b^*, we can also use them when there is no adjustment of the 4 digits level.

In Part 2 of Table 1, the consistency evaluation t able titled "Evaluation of Consistency of Basic Trade Statistics of AID-XT in East Asian Nations and Regions and the US," is represented and the trade data used for making the consistent table is the adjusted data by the way of 4 digit level classification ode correction..