Chapter1　Conversion of Trade Statistics with Reversion to Commodity Classification : Case Study of Japan and Korea

Chapter 1

Conversion of Trade Statistics with Reversion to Commodity Classification:

Case Study of Japan and Korea

NODA Yosuke

The Standard International Trade Classification (SITC) system devised by the United Nations has undergone revision in line with changes made to the Nomenclature of the Customs Cooperation Council (CCC). The sequence of revisions has taken the SITC from Revision 1 (SITC-R1) through Revision 2 (SITC-R2) to Revision 3 (SITC-R3). Consequently, the use of trade statistics as time-series data extending back into the past can encounter difficulty around the times that commodity classifications were revised, because the commodity definitions and their coverage are not necessarily the same. Therefore, it is sometimes not possible to directly use transaction values and quantities for consecutive years that include revision years. However, the United Nations has created tables of correspondences with each revision, to list the correspondences between SITC-R1 and SITC-R2, then between SITC-R2 and SITC-R3, so that revision of the commodity classifications would not make trade statistics unusable.

When there are correspondences between related individual classification codes in Classification *A* and Classification *B* that have different classification systems, as with the correspondences between SITC-R1 and SITC-R2, then the set of closed correspondences between the codes in classifications *A* and *B* are said to be a group. In the case of a commodity classification, this will be a commodity group. When using correspondences that have been grouped, knowledge of the nature of the correspondences between the two classifications becomes an important issue. The combinations of clustered groups of correspondences can be divided into four types, which are type 1, type 2, type 3, and type 4. When considering correspondences with a direction from classification *A* to classification *B*, then the clusters of individual classification codes in *A* and *B* have a one-to-one correspondence in type 1. In type 2, they have a one-to-many correspondence, and in type 3 they have a many-to-one correspondence. Type 4 describes a many-to-many correspondence, and can be further identified as belonging to two types, which are type 4a and type 4b.

When working with correspondences grouped in these ways, the question arises of how statistical values shown under *A* should be converted to corresponding statistical values to be compiled under *B*, and also, conversely, how statistical values shown under *B* should be converted to corresponding statistical values to be compiled under *A*. The subject matter of this chapter is the classification differences that arise due to revision of the commodity classifications used in trade statistics, and an attempt at an application for such differences.

With type 2 and type 4, it becomes necessary under certain conditions to distribute *A* statistical values into *B*, taking account of the weight to be assigned those values. This weight is called the distributed weight and it is a coefficient used in order to take account of the appropriate weight to be assigned when distributing the statistical values under individual classification codes in *A* to the corresponding individual classification codes in *B* for such codes that exist in grouped correspondences. Figure 1 shows the relationship of distributed weight.

In this chapter, we compile trade statistics by creating time-series data for export-import series in SITC-R1, SITC-R2, and SITC-R3 on the basis of the distributed weight that have been obtained. These trade statistics are compiled from the OECD-edited trade statistics for Japan and the UN-edited trade statistics for Korea from 1962 to 1999.

The table of correspondences for basic items that are detailed classification codes in the SITC-R1 and SITC-R2 commodity classifications can be obtained from the *Standard International Trade Classification, Revision* 2 published by the UN Statistics Division. This table of corresponding codes is not necessarily consistent at every point, and problematic portions may be encountered throughout. The Institute of Developing Economies has created a table of corresponding codes that corrects for these problematic points, and has submitted it for use with trade statistics. In modeling the correspondences, we have used the table of corresponding codes corrected by the Institute of Developing Economies as our basic model. Our table of corresponding codes for basic items in the SITC-R2 and SITC-R3 commodity classifications is based on correspondences obtained from the *Standard International Trade Classification, Revision* 3 published by the UN Statistics Division.

The correspondences of the SITC-R1 to the SITC-R2 and the SITC-R2 to the SITC-R3 at the 1 digit level and 2 digits level of SITC are shown under "SITC-R1: SITC-R2" and "SITC-R2: SITC-R3" in Table 2 in Part 2. Correspondences at the 3 digits level of SITC are shown in Table 3 in Part 2, and at the same time this table also shows the distributed weight assigned to the individual classification codes for Japan and Korea. Table 2 in Part 2 and Table 3 in Part 2 show the related information that is considered necessary to model the correspondences. The symbols under the various items of related information and the meanings they represent as follows: $G_i(j)$: this represents a group or sub group, where *i* is the consecutive number of the group and *j* is the consecutive number of the sub group. There are no sub groups in the correspondences of the basic model, so grouped j all have the value 1. *type*: this represents the type of the sub group correspondence. Commodity groups are shown as commodity group i j: type, $G_i(j)$:type, and *Gi-j:type*. The *type* is often omitted. X_A: This represents a classification code in Classification *A*. X_B: this represents a classification code in Classification *B*. X_A-*f*: this represents the frequency of a classification code in A. X_B-f: this represents the frequency of a classification code in B. X_A-Q: this represents the consecutive number of a classification code in A when the codes are arranged in ascending order. X_B-Q: this represents the consecutive number of a classification code in B when the codes are arranged in ascending order. The number of individual classification codes from A contained in the group is defined as m and the number of individual classification codes from B is defined as N.

This chapter takes up conversion that goes in the direction from A to B. These correspondences and the distribution structure of the distributed weight are presented in organized form in Figure 1, where A_i represents the individual classification code from A and x_i represents the statistical value related to the individual classification code A_i. Similarly, B_j represents the individual classification code from B and y_i represents the statistical value related to the individual classification code B_j. If k pieces of data can be obtained each from A_i and B_j as observed values, then the statistical value X from A and the statistical value Y from B will be represented as:

$$X = \begin{bmatrix} x_{11} & \cdots & x_{1k} \\ \\ x_{n1} & \cdots & x_{nk} \end{bmatrix} \quad Y = \begin{bmatrix} y_{11} & \cdots & y_{1k} \\ \\ y_{n1} & \cdots & y_{nk} \end{bmatrix}$$

The weight for distribution in a direction going from A_i to B_j is defined as:

$$\grave{u}_{ij} \quad i = 1 \cdots m,\ j = 1 \cdots n$$

The distribution method involves taking the statistical values x_i corresponding to the individual classification code A and splitting it into n pieces of y_j. The sum of these for y_j is shown in the following formula. As shown in shadowed format in Figure 1, this can be expressed as:

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

Classification B is made up of n elements, $j = 1 \cdots n$. Showing these as a matrix, they are:

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

Taking the weight distribution matrix as W and then saying $\omega_i' = (\omega_{1i} \cdots \omega_{mi})$, we have:

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

If a matrix U is similarly defined, taking k pieces of observed values into account, then the relationship between A, B, and the weight W can be expressed as:

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

If the m-dimensioned vector composed of elements that are all 1 is defined as l_m, then, from the conditions for weight distribution in formula (1), they satisfy the identity:

$$W' l_n = l_m$$

Formula (1) is a form of expression used to model neural networks.

Trade statistics data are generally made up from the six categories of reporting country, the export-import category, the commodity classification, the counterpart country, the year, and the unit of quantity, together with the two statistical values of value and quantity. The trade statistics data used in this chapter are based on data in the Ajiken Indicators on Developing Economies: Extended for Trade Statistics (AID-XT), which is a world trade statistical database system owned by the Institute of Developing Economies. This was created by converting trade statistics compiled by OECD for the Japan data, and trade statistics compiled by the UN for the Korea data, respectively.

The import-export category is shown as 1 for imports and 2 for exports. The year is expressed by the last two digits of the dominical year. IDE unified units of quantity is used for the units of quantity, and the codes and designations of the units of quantity are shown in Table 3. Values are given in units of US$1,000.

The trade statistics for Japan and Korea are estimated at the 3 digits level for each series of the SITC on the basis of distributed weight obtained using the neural network described in Chapter 2 of this book. The estimated values for the distributed weights are shown in Table 3 in Chapter 2. The SITC series created on the basis of these weight are represented as *NN*. The same series created with uniformly distributed weights will be represented as *ED*.

The respective *NN* and *ED* series are estimated for exports and imports under SITC-R1 that relate to Japan. As shown in Table 4, the original series for the period from 1962 to 1977 exists as an SITC-R1 series in Japan, so this was used. For the years 1978 to 1987, an SITC-R2 series exists as the original series. Therefore, a weight distribution with a direction going from SITC-R2 to SITC-R1 was used to estimate an SITC-R1 series. The arrow symbol indicates the direction of the conversion. For the years 1988 to 1999, the SITC-R3 exists as the original series. Therefore, the SITC-R3 was used to estimate an SITC-R2 series, and then the SITC-R2 series this yielded was converted to an SITC-R1 series. The time-series data for the SITC-R1 series in the 1962–1999 range can be obtained by putting together these converted series. SITC-R2 and SITC-R3 series can also be derived by conversion in a similar manner.

For Korea, it was not possible to estimate the distributed weight for commodity group 61 in the correspondences between SITC-R2 and SITC-R3. As a consequence, no time-series data were created for the period from 1962 to 1999. Instead, the respective series for SITC-R2 were created in the same manner as the *NN* and *ED* series for exports and imports under SITC-R1 for the period 1962 to 1987. The SITC-R3 series has been kept just as it was in the original series.