Chapter 2

Evaluation and Correction for Consistency of Trade Data:

Case of Japan and Korea

NODA Yosuke

This chapter will discuss, using concrete examples, the methods used to evaluate and correct the consistency of data for imports to Japan and South Korea as reporting countries in UN Comtrade data, based on Chapter 1, "Evaluation and Correction for Consistency of UN Comtrade Trade Data." Chapter 1 discussed methods of evaluating the consistency of trade data and of correcting inconsistent data, but the discussion in that chapter was predicated on the conditions that data for partner countries was consistent, and that the digit-level classification codes were all identical for units of quantity. Actual trade data does not necessarily satisfy these conditions. Differences in the definition or method of formulation of trade data are directly reflected in UN Comtrade data, and the conditions mentioned above take a variety of forms in response to these differences.

When the reporting country, year, and direction of trade are given, the error in a trade matrix formulated on the basis of hierarchically organized digit-level commodity classifications and individual partner countries, using commodity totals and transaction values for partner country world as benchmarks, can be divided into commodity classification-related error, partner country-related error, and total error, which includes commodity classification-related error and partner country-related error. The sum of these errors is total error. Consistency evaluation tables are formulated by expressing these errors. Another method of evaluation of consistency is to study the difference between the transaction values for the sums of commodity classified according to a higher-level classification code and a lower-level code for which the higher-level code is the next level. Based on these two types of consistency evaluation table, trade data is corrected to maintain consistency for partner countries and commodity classification codes. The correction criteria are determined on the basis of absolute error and relative error. Focusing on imports for Japan and South Korea as reporting countries, this chapter looks at the commodity classifications used in trade data, the evaluation of the consistency of trade data, the correction of inconsistencies in trade data, and a variety of consistency evaluation tables. The commodity classifications for both reporting countries are SITC-R1. The fiscal years covered by the data are 1962-2005 for Japan, and 1962-2005 for South Korea. In this chapter, the correction criterion for absolute error, α*, in equation (4-2) discussed in Chapter 1 is 1 US$.

1. Evaluation for Consistency of Trade Data

Appendix 1 is a consistency evaluation table for SITC-R1 digit-level classification codes for imports to Japan (1962-2005). The appendix was formulated with the correction criterion for absolute error, α^*, set at 1. The appendix shows that if α^* is set at 3,000 US$, the total error, e, in a consistency evaluation from 1

digit-level classification codes to 3 digit-level classification codes for Japan as reporting country is low, and consistency is maintained for both commodity classifications and partner countries. For 4 digit-level classification codes, a lack of consistency of 2% or higher is observable for relative error from 2002 onwards. However, because consistency is maintained with regard to error arising from partner countries, these errors can all be judged as arising from commodity classifications. From Figure 1, the fiscal years for which error occurs for Japan are the years covered by the HS2002 classifications, and the errors in 4 digit-level classification codes can be considered to be errors arising from conversion from HS2002 to SITC-R1.

The fiscal years in which the highest values for total error occur for the most detailed classification codes (*mdcc*) in Table 1 are 1962, 1963, 1965, 1967, and 1988 onwards; these errors can also all be judged as originating in product classifications. Before 2001, no error occurs in 4 digit-level classification codes, and it is therefore assumed that error occurs between 4 digit-level and 5 digit-level classification codes. From 2002 onwards, there are no changes in error between items, and error is therefore assumed to originate between 3 digit-level and 4 digit-level classification codes.

Table 5 shows an evaluation of the consistency of digit-level classification codes with Japanese imports for 1977 as an example. Table 8 shows the consistency of transaction values for digit-level classification codes and the sums of transaction values for lower level classification codes. -848 in the error originating in product classifications for 1 digit-level classification codes in Table 5 matches the error between digit-level classification codes, $\alpha(c_{\bullet}(0), W)$, in Table 8. For classification code {m} as the item for correction in the 1 digit-level classification codes, assuming

$v(m, W) = \alpha(c_{\bullet}(0)) = -848$; using this in combination with the original trade data, correction can be applied by formulating Equation (4-5) from Chapter 1. Table 9 shows the results of correction. The total transaction values for 1 digit-level classification codes including the corrected item, {m}, matches the total value for products, and the error for digit-level classification codes becomes 0.

Table 5 shows the consistency between transaction values for digit-level classification codes for Japanese imports (1962-2005) and the sums of transaction values for lower-level codes. The units for transaction value are 1,000 US$. The error of 7,997 (the units are 1,000 US$) originating in product classifications for 1962 as shown in the *mdcc* in Table 1 is shown as the error originating between 4 digit-level and 5 digit-level classification codes in Table 5; there is a difference of 7,997 between the transaction value for classification code {2839} in the former and the sum of the transaction values for 28391 and 28392 which are elements of $D_S(2839)$ in the latter. In relation to the consistency evaluation table, $e_c(4) + e_{c,p}(4)$ is 0 (the units are 1,000 US$), and $\alpha(2838)$ and $e_c(5) + e_{c,p}(5)$ therefore match and make 7,997. The transaction value of $D_S(2839)$ in the breakdown of 5 digit-level classification codes in Table 5 is shown as q. Table 3 shows a consistency evaluation for SITC-R1 digit-level classification codes for imports to South Korea for 1962-2006. Table 3 is formulated with the correction criterion α^* set at 1. If α^* is set at 3,000 US$, low-level error is generated in total error for 1 digit-level classification codes for several fiscal years. Error is generated in the same fiscal year and other fiscal years for classification codes from the 2 digit to the 4 digit level. The highest level of error is concentrated in the period from 1988 to 1998, a period in which HS1988 and HS1996 classifications were employed in South Korea. Total error is thought to originate in the conversion from these HS revisions to

SITC-R1. Error also occurs in 4 digit-level classification codes from 2001 onwards, but this is all error originating in product classifications. For *mdcc*, the fiscal years in which the highest level of total error occurs are each fiscal year from 1962 to 1967 and from 1988 onwards, and this error all originates in product classifications. Table 6 shows the consistency of the transaction values for digit-level classification codes and the sum of transaction values for lower level codes for South Korean imports for 1992.

2. Correction of Trade Data

This section will provide an example of the method used to correct data for Japan. Error for Japanese imports appears as the inconsistency in 4 digit-level classification codes for 2002 onwards in Table 1. Taking 2004 from Table 5 as an example, $D_S(332)$ and $D_S(641)$ show error for 2004; the total of these is 10,909,055, a figure which matches the error for 4 digit-level classification codes. Correction was applied to add 332m and 641m, and the former was formulated as $D_S(332)^*$ and the latter as $D_S(641)^*$. For 2004, the corrected items were 332m and 641m for 4 digit-level classification codes. Table 2 is the consistency evaluation table (1962-2005) formulated using the corrected *mdcc* classifications for Japanese imports.

Correction was applied in the same way for South Korea. This will be discussed using 1992, for which total error is at its maximum in 3 digit-level classification codes in Table 3, as an example. Table 3 shows that total error and error relating to classification

codes is relatively high for 1992 for 2 digit-level to 4 digit-level classification codes. Error between digit-level classification codes in Table 6 was an item for correction. d_3, d_4, and d_5 in the *mdcc* for 1992 in Table 3 are 0, 348, and 797 respectively. By contrast, in Table 4, the consistency evaluation table for *mdcc* classifications based on corrected trade data, they are 16, 357, and 798 respectively. The result of subtracting the number of product classifications in Table 3 from the number of product classifications in Table 4 is the number of classifications added for correction. As a result, 16, 9, and 1 classifications were added to the 3 digit-level, 4 digit-level, and 5 digit-level classification codes respectively. Table 4 is the consistency evaluation table for 1962-2006 based on corrected *mdcc* for South Korean imports.

In this chapter, the correction criterion for absolute error, α^*, was set at 1 US$. However, changing the correction criterion for absolute error also changes the total error as it relates to consistency, and this changes the number of classification codes which it is necessary to correct. Table 10 shows a consistency evaluation for corrected *mdcc* for Japanese imports for 1977, obtained by changing the evaluation criterion, α^*. When the evaluation criterion is set at 0, total error is 0, and the number of digit-level classification codes for corrected *mdcc* is 1, 9, 32, 458, and 950 for the 1 to 5 digit levels respectively. When the evaluation criterion is set at 1, total error remains at 0; the number of digit-level classification codes remains the same from the 1 to 3 digit levels, but the number declines at the 4 and 5 digit levels.