# Preface

The system of commodity classification for international comparisons used in trade statistics is basically comprised of two trade commodity systems. One is Standard International Trade Classification (SITC) prepared by the United Nations (UN). This classification system includes the SITC (original version), its revision (also called Revision 1 and abbreviated as SITC-R1), Revision 2 (SITC-R2), and Revision 3 (SITC-R3). The other is the Customs Cooperation Council Nomenclature (CCCN) prepared by the Customs Cooperation Council (CCC). This system started from the 1955 Brussels Tariff Nomenclature (BTN), the first commodity index, and subsequently evolved into the CCCN and then to the current Harmonized Commodity Description and Coding System (HS). The HS classification system includes the first version of HS based in 1988 (HS88 version also called the HS-R1) and the subsequent revisions to the 1996 revised HS96 (HS-R2). Correspondence between these two classification systems has been established, though comparatively rough in state, such as with the CCCN and SITC-R2, and the HS and SITC-R3.

Even in the same SITC classification system, there is not necessarily any consistency or match in correspondence before and after revisions. Sufficient care must be taken in the use of time-series data of trade statistics. The SITC-R2, the revision of the SITC-R1, was considerably revised in the most detailed individual classification codes, the basic commodity items. Further, several changes and modifications etc. have been made even at the higher levels. Basically, however, a structure of commodity classification system the same as the SITC-R1 has been maintained. The SITC-R3, the revision of the SITC-R2, however, has a vastly different commodity classification system compared with the SITC-R2 even in the concept of classification.

As an attempt for applying time-series trade statistics edited from commodity classifications of different classification systems, Noda and Yamamoto proposed, for correspondence between the SITC-R2 and the SITC-R3, the idea of establishing correspondence between commodity groups not by correspondence of basic items of the commodity classification, but by grouping a number of basic items together. Further, if the portion assumed not to have much relationship in the correspondences existing in the commodity groups are excluded, the commodity groups are said to be "cut" when commodity groups are divided into several commodity sub groups.

The commodity groups and the commodity sub groups are determined by the method of cutting, so the method of cutting can be considered the model for grouping correspondence codes. Grouping and cutting the commodity classifications is, in other words, to model correspondences by linking all the related correspondences between basic items of the SITC-R2 and SITC-R3 of the different systems to form commodity groups, and cutting them to form sub groups.

The model of correspondence is prepared with the aim of reducing as much as possible the mis-

match of the trade statistics caused by the conversion in connection with revisions of commodity classification. If the correspondence becomes complicated such as seen by the correspondence between the SITC-R2 and SITC-R3, however, with this method, there is the problem that the commodity groups become too large and it becomes impossible to identify the corresponding commodity group in actual use. Therefore, it becomes necessary to estimate the weight to be distributed under certain conditions to the individual classification codes when extracting commodity groups as sets of individual classification codes.

Recently, the method of data mining for obtaining knowledge from the large amount of data accumulated due to the great strides made in the processing capabilities of computers has been developed and put into practical use. This book attempts direct conversion of trade statistics by individual classification codes by applying the method of data mining by neural networking to estimate the distributed weight in the correspondence of commodity classifications. Converted data is trade statistics with export and import from 1962 to 1999, of Japan of the trade statistics compiled by the OECD, and of The Republic of Korea of the trade statistics compiled by the UN.

To use trade statistics of annual data available for only a small number of years as data for the technique of neural networking, which inherently is based on a large amount of data, considerably strict hypotheses have to be set. The same commodity classifications are used without consideration of the import and export classifications, at the 3 digits level of the SITC. Further, to eliminate the fluctuations inherent to economic data occurring due to economic fluctuations including long term trends

and cycles, the amounts of transactions are not used directly. Instead, ratios between commodity groups are used as data.

While the results of the current study revealed that there are several problems which must be dealt with, it is believed that neural networking can be effectively applied to the conversion of trade statistics accompanying revisions in the commodity classification. Due in part to the restrictions in the amount of data, further problems remain to be studied including how to ease the strict hypotheses.

This book is comprised of four chapters and reference materials. Chapter 1 "Conversion of Trade Statistics with Revision to Commodity Classifications: Case Studies of Japan and Korea" provides a summary of the entire book. This chapter summarizes the methodology for estimating the series from Revision 1 to 3 on the commodity classification SITC 3 digits level of Japan and South Korea. Due to space limitations, it is impossible to list the trade statistics of all trade partners for each SITC 3 digits level, but the data can be provided by CD-ROM. The methodology used in this chapter is explained in detail in Chapter 2 to Chapter 4.

Chapter 2 "Estimates of Weight Distribution to Conversin for 3 Digits Level of SITC: Application of Neural Network Method" briefly explains the method of data mining by neural networking and specific method of estimation of the distributed weight. This method has been interpreted negatively as a "blind search for data without prior hypothesis for verification", but recently it has begun to be used in various fields as methodology for data processing as its effectiveness has been shown. Of course, this is the first attempt to use this method for conversion of trade statistics in connection with the revision of commodity classifications.

Chapter 3 is entitled the "Evaluation of Consistency of Trade Statistics with Revision of Commodity Classifications". It studies if there is a fault in data due to an inconsistency of classifications at the time of revision of the commodity classifications in the obtained SITC series. It is not necessarily possible to determine whether the fault is due to an inconsistency caused by the revision, but the position is taken that it is possible. Since the existence of a mismatch is treated as a fault, that is, a changing point, the regression model and Bayes-type posterior distribution are applied as the methodology for the structural changes due to the time series. It will have to be decided in the future whether the cause of commodity groups being recognized as a changing point is due to the correspondence or due to the estimation of the distributed weight. Suffice to say that in using the data, it is necessary to take care in handling the relevant commodity groups.

Chapter 4 is entitled "Application of Country Codes in Database of Trade Statistics". In the same way as revision of commodity classifications, the problem in long-term time series use of trade statistics is the changes in countries and customs areas. This chapter throws light on the relationship between the UN Trade Statistics, the OECD Trade Statistics, and the Taiwanese Trade Statistics owned by the IDE and the "IDE Standardized Country Code" prepared for international comparison by the IDE. Further, it studies the method for coding and retrieving the countries and costoms area in the database of trade statistics.