Chapter 3

Estimation of Distributed Weight Matrix for Common Commodity Classification and its Transformation

NODA Yosuke

The UN COMTRADE database contains data in which post-revision commodity classification codes have been converted into pre-revision commodity classification codes in order to enable the data to be employed in long-term time series using pre-revision classifications. The method employed by the UN to do so is, as the UN itself acknowledges, a rough estimation method. It treats data uniformly without consideration of the distributed structure based on the correspondence tables formulated to enable conversion, and without consideration of the differences due to reporting country and whether the data relates to imports or exports. The method employed by the IDE, by contrast, involves estimating distributed weight matrices that consider the distributed structure of the relationships of correspondence within each commodity group, and the transaction value for each reporting country and import/export category.

The distributed weight matrix acts as a filter when estimating the transaction value for a pre-revision commodity classification, B, by converting the transaction value for a post-revision commodity classification, A, using the relationship of correspondence from A to B. The matrices are estimated using the value of transactions for commodity classifications pre- and post-revision. The hypothesis discussed below is essential to enabling the estimation of the matrices. It is on the basis of this hypothesis that the distributed structure is formulated that enables the transaction value for classification A to be converted into the transaction value for classification B on the basis of the distributed weight matrices from A to B.

When a relationship of correspondence from classification A to classification B exists for a commodity group, there is assumed to be no major change in the structure of transaction value from year to year for the classification either pre- or post-revision. A sample taken at random from the pre-revision period when the structure of the transaction value of classification A is stable is interpreted as the constituent ratio of the transaction value for that classification. The constituent ratio of classification B is similarly assumed to be represented by a sample taken at random when the structure of the transaction value for the classification is stable. Samples are assumed to be taken simultaneously from both classifications for this period.

A sample corresponding to the constituent ratio of the transaction value of both classifications cannot be obtained from the same period. Separate constituent ratios are obtained from the pre- and the post-revision periods. The hypothesis that the sample obtained from the pre-revision period can be obtained using the same random sampling as employed to obtain the post-revision sample may be seen as rather bold. The important point here is that the structure of pre-revision classification *B* is maintained unchanged post-revision, and that the sample obtained is treated not as the transaction value, but as the constituent ratio that expresses the structure of the transaction value. The respective transaction values $x_1^D \cdots x_n^D$, are assumed to correspond to the *n* individual classification codes of classification *A* in the product groups, $a_1 \cdots a_n$. For $j = 1 \cdots n$, x_j is expressed as the vector made up of *h* samples corresponding to annual data, $x_j^D = (x_{j1}^D \cdots x_{jh}^D)'$ and

$$\begin{pmatrix} x_1^{D}{}' \\ \vdots \\ x_n^{D}{}' \end{pmatrix} = \begin{pmatrix} x_{11}^D & \cdots & x_{1h}^D \\ & & \\ x_{n1}^D & \cdots & x_{nh}^D \end{pmatrix} = X^D$$

If the transaction values for individual classification codes of classification *B* ($b_1 \cdots b_m$) are expressed as $(y_1^D \cdots y_m^D)' = Y^D$.

As given above, the condition that all the transaction values for classification *A* in the same year are converted by means of distributed weights into transaction values for classification *B*, which is the implication of the hypothesis that the transaction values for classifications *A* and *B* for the same year are simultaneously distributed. The total of transaction values for classification *A*, $(x_{1j}^D \cdots x_{nj}^D)$, is allocated with respect to *j*, which represents the year in conversions from classification *A* to *B*, and must match the total for classification *B*, $(y_{1j}^D \cdots y_{mj}^D)$. If $h = k$, the sum of the transaction values for each year match, and

$$(x_{\bullet 1}^D \cdots x_{\bullet k}^D) = (y_{\bullet 1}^D \cdots y_{\bullet k}^D)$$

Because the transaction values for classification *A* are maintained without change in classification *B*, $x_\bullet^D = y_\bullet^D$ for the entire commodity group. That is, because $y_{\bullet j}^D = x_{\bullet j}^D$ should take a fixed value for $j = 1 \cdots k$, this constant has been set as 1. Taking the constituent ratios of the respective matrices for the transaction values of classifications *A* and *B*, (matrices *X* and *Y*), satisfies this condition. $D(x)$ is the diagonal matrix created treating vector *x* as a diagonal element. If $X = X^D D(l_m' X^D)^{-1}$ a matrix that has constituent ratios as elements can be formulated. The same holds true for Y^D.

1. Structure of Distributed Weight Matrix

Embed Equation.3 ω_{ij} is the distributed weight from classification codes a_j to b_i when converting from classification *A* to *B* in commodity groups. If $\omega_{ij} \neq 0$, y_i' for b_i in classification *B* can be expressed as

$$y_i' = x_1'\omega_{i1} + \cdots + x_n'\omega_{in} + u_i'$$

n relation to the relationship of correspondence with the distributed weight for classification *A*, $i = 1 \cdots m$. For $j = 1 \cdots n$, $\omega_{1j} + \cdots + \omega_{mj} = 1$, and u_i is the disruption term in vectors possessing the same structure as y_i. If the distributed weight matrix for the *m* ×*n* matrix in which there is complete correspondence from classification *A* to *B* is termed *W* and the characteristics of the distributed weight matrix satisfy the weight condition $l_m' W = l_n'$. Expressing the transaction values for classifications *A* and *B* (*X* and *Y*) and the distributed weight matrix *W* as a matrix gives

$$\begin{pmatrix} y_1' \\ \vdots \\ y_m' \end{pmatrix} = \begin{pmatrix} \omega_{11} & \cdots & \omega_{1n} \\ \vdots & & \vdots \\ \omega_{m1} & \cdots & \omega_{mn} \end{pmatrix} \begin{pmatrix} x_1' \\ \vdots \\ x_n' \end{pmatrix} + \begin{pmatrix} u_1' \\ \vdots \\ u_m' \end{pmatrix}$$

and this can be expressed as

(1-1) $$Y = WX + U$$

Due to considerations of space, the discussion in this chapter will not consider the type of relationship of correspondence, the least squares method with equality constraints, or the entropy optimization method.

2. Estimation of *W* based on Contingency Tables

Substituting *U*=0 in Equation (1-1), which expresses the structure of the distributed weight, and multiplying both sides of the equation from the right by l_k/k gives $(\bar{y}_1 \cdots \bar{y}_m)' = W(\bar{x}_1 \cdots \bar{x}_n)' = WD(\bar{x})l_n$
When *Q* is an appropriate figure,

(2-1) $$V = WD(\bar{x})Q$$

and if the respective elements of *V* are assumed to be integers, then Equation (2-1) is the distributed value matrix expressed by $\bar{x}$. If the sums of the outer columns and rows of *V* are $l_m'V = \bar{x}'Q$ and $Vl_n = \bar{y}Q$ respectively, then the sum of *V* is $l_m'Vl_n = \bar{x}'l_nQ = l_m'\bar{y}Q = Q$.

Treating the distributed value matrix *V* as a two-dimensional contingency table, when the respective elements of *V* are considered as stochastically distributed random variables, *V* can be postulated as a contingency table distributed according to polynomial distribution of joint probability functions. If the achieved value of V_{ij} for random variables $i = 1 \cdots m$ and $j = 1 \cdots n$ is termed v_{ij}, and $P\{V_{ij} = v_{ij}\} = p_{ij}$, the joint probability function can be expressed as

$$f(V_{11} = v_{11} \cdots V_{mn} = v_{mn}; p_{11} \cdots p_{mn}) = Q!(\prod_{ij} v_{ij}!)^{-1} \prod_{ij} p_{ij}{}^{v_{ij}}$$

Naturally, $p_{\bullet\bullet} = 1$. Treating *a* as a constant term unrelated to p_{ij} gives

(2-2) $$\ell(p_{11} \cdots p_{mn}) = \log f = a + \sum_{ij} v_{ij} \log p_{ij}$$

for the log likelihood function. In addition, because the total value of *V*, *Q*, has been determined, it is possible to express the transaction value represented by the respective elements as

(2-3) $$v_{ij} = p_{ij}Q = p_{\bullet j}\, p_{i|j}Q$$

Here, $p_{\bullet j}$ is the marginal probability of a_j in classification *A*, and $p_{i|j}$ is the conditional probability of b_i in classification *B* when the probability of a_j is known.

We assume that relationships of correspondence exist between each individual commodity code in classifications *A* and *B*. Terming the marginal probability of b_i in classification *B* $p_{i\bullet}$, $p_{i|j} = p_{i\bullet}$ in Equation (2-3) when the joint probabilities are independent. Given this, v_{ij} can be expressed as $v_{ij} = p_{ij}Q = p_{\bullet j}\, p_{i\bullet}Q$. Therefore, if the joint probability matrix of which p_{ij} forms an element is termed *P*, $P = D(Pl_n)l_m l_n' D(l_m'P)$. In addition, if v_{ij} is expressed as a matrix,

(2-4) $$V = D(Pl_n)l_m l_n' D(l_m'P)Q$$

When *V* is given, *W* can be calculated. The distributed weight matrix is given by

(2-5) $$W = V\{D(\bar{x})Q\}^{-1} = D(Pl_n)l_m l_n'$$

and can be calculated using only the marginal distribution of classification *B*.

When classifications *A* and *B* are mutually independent, $\hat{p}_{i\bullet}$ and $\hat{p}_{\bullet j}$, the solutions that maximize the log likelihood function (Equation 2-2), are the maximum likelihood estimators for $p_{i\bullet}$ and $p_{\bullet j}$ respectively. The Lagrange function with $p_{i\bullet}$ and $p_{\bullet j}$ as constraint conditions is expressed as

$$s = \ell(p_{1\bullet} \cdots p_{m\bullet}, p_{\bullet 1} \cdots p_{\bullet n}) + \mu\left(\sum_i p_{i\bullet} - 1\right) + \eta\left(\sum_j p_{\bullet j} - 1\right)$$

The maximum likelihood estimators are found from $\hat{p}_{i\bullet} = v_{i\bullet}/Q$ and $\hat{p}_{\bullet j} = v_{\bullet j}/Q$, and the maximum likelihood estimator for the transaction value matrix is $\hat{V} = D(\bar{y})l_m l_n' D(\bar{x})Q$. The maximum likelihood estimator of the distributed weight matrix given by Equation (4-9) is $\hat{W} = D(\bar{y})l_m l_n'$. Substituting $W_2 = D(\bar{y})a(W)$ and reformulating the equation to satisfy weight conditions gives

$$W_i = W_2 \cdot D(l_m{}' W_2)^{-1} \tag{2-6}$$

The method of formulating distributed weight matrices with identical distribution patterns, based on the assumption of the independence between the classifications, is termed the *i* method.

3. Characteristics of Methods of Formulation

When the transaction value constituent ratio, *X*, of classification *A* has been determined and the true value of the general distributed weight matrix *W* is known, Equation (1-1), which expresses the structure of the distributed weights, can be used to calculate the transaction value constituent ratio, *Y*, of classification *B*. Working in reverse order, *W* can be found from *X* and *Y* by imparting error to *U* in Equation (1-1) and formulating *Y*. Because the value of *W* is already known, it can be used as a standard in determining the accuracy of *W* as calculated by different calculation methods in terms of change in the degree of error.

As calculation methods, (1) the method of simple averages (termed *s*), (2) the identical distribution pattern formula based on the assumption of independence between the classifications (termed *i*), (3) the UN method, in which the maximum value of the particular solutions of the *i* method is termed 1 and the others 0, and (4) the method of determining directly the distributed weight matrix using the least squares method with equality constraints (termed *wm*), were selected for comparison using the transaction value X^D and the true distributed weight matrix, *W*, which represent the actual situation more accurately than the transaction value constituent ratio. In addition, the distributed weight matrices formulated using methods (1) and (4) were used as initial values for entropy optimization. These are termed *s2* and *wm2* respectively.

In calculation method *wm*, the true value of *W* is calculated when Y^D displays no error, but as the magnitude of error increases, the results show considerable fluctuation around the true value. Because *W* is formulated based on the total for *Y* in the *i* method, even when *Y* displays considerable error, the calculated value does not differ significantly from the calculated value when *Y* displays no error. *wm2*, in which the entropy optimization method was applied to the results of *wm*, displays the same characteristic as *wm*; as the error of Y^D increases, results display greater variation against the true value. Attention must be paid to the fact that *i2* and *s2* result in the same values.

When the true value of the distributed weight matrix is known, the least squares method with equality constraints reacts sensitively to error in the structure determined by Equation (1-1), i.e. the error in Y^D for the transaction values corresponding to classification *B*. By contrast, methods *i*, *i2* and *s2* are not sensitive to error, and produce largely constant results, which may, however, not be close to the true value.