A Documentation on Hong Kong’s Re-export Markup Estimation

Shunli Yao

April 1999

1. Introduction

This serves as a documentation for the work on Hong Kong’s re-export markups in Feenstra, Hai, Woo and Yao (1998, 1999). It also prepares SAS programs that extend the analysis to countries other than the US.

2. The Data

The data used in this work are Hong Kong’s import and re-export data and China’s export data. The Hong Kong data was acquired from the Census and Statistical Department, Hong Kong Special Administrative Region. The China data was provided by China’s Customs General Administration.

Hong Kong Data

The Hong Kong data covers the years 1984-97. The 1984-91 data is classified by 6-digit SITC (rev. 2) codes; the 1992-93 data is classified by 5-digit SITC (rev. 3); and the 1994-97 data is classified by 8-digit HS codes.

The original data came with annual and current month (December) data. The annual data was later singled out as separate files and stored under the directory “/export/black-home2b/yao/home6/yao/hk8497”. The annual datasets are in SAS format. For Hong Kong import data, the file names are: imp**y.ssd04; for Hong Kong re-export data, the file names are: rep**y.ssd04; and for Hong Kong export data, which is not used in the markup calculation, the file names are: exp**y.ssd04; where **=84-97 and 'y' stands for yearly data.

The 1984-91 import data contains the following variables: (1). The commodity codes (SITC_R2); (2). consignment country codes (CTY_C); (3). the year of the data (YEAR); (4). import value (IVALUE_Y); and (5). import quantity (IQUANT_Y). For the import data, the consignment countries refer to the origin countries. The import value is in Hong Kong dollar. The 1992-97 import data has similar structure as 1984-91 data, except that the commodity codes are SITC_R3 (for 1992-93) and HS for 1994-97, and there is an additional variable that specifies the level of commodity classifications (TRANS), whereas TRAN='1’ refers to the most disaggregate data.

The re-export data contains the following variables: (1). The commodity codes (SITC_R2 for 1984-91 data, SITC_R3 for 1992-92 data and HS for 1994-97 data); (2) codes for origin countries (CTY_O); (3). Codes for consignment countries (CTY_C), which refer to the destination country codes; (4). Re-export value (RVALUE_Y) in Hong Kong
dollar; (5). Re-export quantity (RQUANT_Y); and (6). The year (YEAR). For 1992-97 data, there is a variable TRAN that describes the transaction type of the data, i.e., the level of data aggregation: for the most disaggregate data, TRAN='1' and the re-export quantity is not zero in general.

China Data

The China data is stored under the directory

```
/export/black-home2b/yao/home6/yao/china
```

The original data is given the filenames “cn**i.ssd04” for import data and “cn**e.ssd04” for export data, where ** represents years (88-96), “i” stands for “import” and “e” stands for “export”. For a detailed documentation of China data, see “China External Trade Data, 1988-1996” archived at IGA.

Since only export data is used in markup calculations, I will focus on the relevant variables in the export files. The data we have is aggregated from the most detailed data originally collected by the China’s Customs. For example, the China Customs collects data for recent years at 8-digit HS code level, while the data we have for recent years is at 6-digit HS levels. During the aggregation process, mistakes were made that for the same commodity code, there are more than one units associated with this code. This problem was fixed by assigning these records a new unit “00” and zero quantity. Another problem with China’s 1988-91 export data is that some SITC5 codes have blank last 1 or 2 digits and some have ‘A” in the third digit. The blank and the third digit with ‘A’ is filled in with ‘0’. The fixed data files were given new names “fcn**e.ssd04”, where “f” stands for “fixed”. It is the fixed export data that was used in markup calculations. Similar procedure is also applied to the import data to produce the fixed data “fcn**i.ssd04”. The program used to fix these problems are fix**.sas, **=88-96, both import and export data (Appendix A). The concordances used for the SITC5 correction are sitc**i0.ssd04 for import data, and sitc**e0.ssd04 for export data, **=88-91.

The fixed export data contains variables relevant to markup calculations: (1). The commodity codes (SITC5 for 1988-91 data and HS6 for 1992-96 data); (2). Countries of origin/destination (CTY_OD: for import data, it refers to country of origin and for export data, it refers to country of destination); (3). Countries of purchase/sale (CTY_PS); (4) value (VALUE); and (5) quantity (QUANTITY). In the markup calculation, only those records with CTY_PS='110' is relevant.

Concordances

Since the data for the same years but from different sources is classified with different commodity codes and country codes, concordances are required to merge the data. Concordances used in the markup calculations are: (1). Concordance between 6-digit HS codes and 5-digit SITC codes with a filename “HS6_SITC5.ssd04” located under directory “/export/black-home2b/yao/home6/yao/china”; and (2). Concordance between
3. Methods of Markup Calculations

Definitions

There are two ways to define the markups: value added to the re-exports as a percentage of import value \((M_1)\); and value added to the re-exports as a percentage of re-export value \((M_2)\). The relationship between the two definitions is: \(M_2 = M_1/(1+M_1)\), which holds at both aggregate and disaggregate levels with proper choice of weights which will be discussed below.

The aggregate markup calculations can be done either at aggregate level, i.e., to calculate the total value added to the re-export as a percentage of the total imports/re-exports; or at disaggregate commodity level first and then calculate their weighted average.

Choice of Weights

If the markups are calculated first at commodity level, the choice of weight depends on the definitions. Let the unit-value of Hong Kong re-export be denoted by \(PM_i = VM_i/QM_i\) where \(VM_i\) is the value and \(QM_i\) is the quantity of imports, and \(i\) denotes the SITC or HS codes. Let the unit-value of Hong Kong re-exports be denoted by \(PX_i = VX_i/QX_i\), where \(VX_i\) is the value and \(QX_i\) is the quantity of re-exports. Thus the relationship between the aggregate markup and disaggregate markup can be shown by the following formulas,

\[
M_1 = \frac{\frac{PM_i}{QX_i}}{\frac{PM_i}{QX_i}} = \frac{PM_i}{QX_i} \times \frac{QX_i}{PM_i} \times M_1
\]

\[
M_2 = \frac{\frac{PM_i}{QX_i}}{\frac{PM_i}{QX_i}} = \frac{PM_i}{QX_i} \times \frac{QX_i}{PM_i} \times M_1
\]

In other words, if the markup is defined as \(M_1\) and is calculated first at the commodity level, the proper weight should be the import unit-value times the re-export quantity; if the markup is defined as \(M_2\) and is calculated first at the commodity level, the proper weight should be the re-export unit value times the re-export quantity.

Markup Presentations
There are three approaches to present the markups, each having different economic interpretations.

Method A: import unit value \( PM_i \) is the origin country specific only, while the re-export unit value \( PX_i \) is both origin and destination country specific. Since Hong Kong’s import data contains imports for re-export and import for local use, and it has only origin country (but no destination country) information, the import unit value computed with Hong Kong’s import data thus represents the unit-value averaged over destination country. The re-export data and the re-export unit value computed with the re-export data, on the other hand, is both origin and destination country specific. The markup computed with Hong Kong’s import and re-export data represents the percentage increase in value of re-exports to a specific destination country relative to the unit value of imports from an origin country to the world.

Based on method A, we can obtain another expression of the markup. The markups at commodity level can be averaged with proper weights discussed above over the destination countries. This markup measure is called method B. It represents the percentage increase in unit-value of re-exports to the world (Hong Kong not included) relative to the unit value of imports from an origin country to the world (Hong Kong included).

The above two methods apply to Hong Kong’s re-export of goods from all countries, using solely the Hong Kong data. With the available China’s export data, the markup calculation can be improved for goods originated from China. Based on method A, the improvement is made to add the destination country information to the unit value of imports from China. In method C, let \( PCTY_C \) and \( QCTY_C \) denote the unit value and quantity of Chinese goods that are exported to a destination country \( CTY_C \) via Hong Kong, and let \( PHK \) and \( QHK \) denote the unit value and quantity of all Chinese exports via Hong Kong within this SITC or HS category \( i \). In general, the two unit value \( PCTY_C \) and \( PHK \) are different, while the latter represents the average unit value purchased by Hong Kong, then re EXPORTed to the world or retained locally. When the Hong Kong import data and China export data is merged, the following formula is used to adjust the unit value of Hong Kong’s import destined for country \( CTY_C \),

\[
PM_i^* \times \frac{QCTY \times C_i}{3 \times QX_i} \times \frac{PCTY \times C_i}{3 \times PHK_i} \times \frac{1 \times PM_i}{QX_i} \times (1 \times \frac{QCTY \times C_i}{QX_i}) \times PM_i
\]

Assign \( QCTY \times C_i/QX_i=1 \) if \( QCTY \times C_i>QX_i \). It can be simplified as

\[
PM_i^* \times COEF \times C_i \times PM_i
\]

where

\[
COEF \times C_i \times 1 \times \frac{QCTY \times C_i}{3 \times QX_i} \times \frac{PCTY \times C_i}{3 \times PHK_i} \times \frac{1 \times PM_i}{QX_i}
\]
Method C markup can be obtained by replacing PM\(_i\) with PM\(_i^*\) in two markup definitions discussed earlier, or it can be calculated from method A markup with the following formula,

\[
M_{2_{2i}}^2 \equiv (M_{2_{2i}}^1 - 1) \times \text{COEF}_C \times 1
\]

This is the formula we used to compute the method C markups in SAS implementation.

**Outliers Detection**

The markup calculation is sensitive to some records with measurement errors. The errors may be caused by re-export time lag, local retention of import goods, or deliberate false report, etc. One way to detect the records with measurement errors is to compare the total re-export quantity with the total import of a commodity. The ratio between the two quantities should be no more than 1, because Hong Kong can not re-export more than it actually imports in the same calendar year. If this ratio is significantly greater than 1, the measurement errors occur for sure. If this ratio is too small, we believe much of the goods imported are for local use and not for re-export. Prior surveys put the average magnitude of markups at 22-29\% by definition M\(_1\), which has an upper limit 1. The distribution of 1988 markups (method A as discussed below) over this quantity ratio (Appendix B) shows, however that, quite a few markups are negative or above 0.5. It shows that for markups grouped by this total quantity ratio, high quantity ratios are associated with high occurrence of negative markups, while low quantity ratios are associated with high occurrence of high markups (>0.5). The same patterns also exist in markups of other years.

To eliminate the measurement errors, we delete the records with the quantity ratio greater than 2. Since the re-export quantity QX\(_i\) is in the weights used to compute the aggregate markups, this aggregation method already has a built-in mechanism to minimize the influence of measurement errors caused by high local retention (low quantity ratio), and thus we do not need to explicitly delete records with low quantity ratio for the average markup calculation.

4. **SAS Implementation**

Since Method B markup is simply the weighted average of Method A over destination countries, SAS programs presented here focus on the Method A and C, which give the commodity level markup calculations. Based on these calculations, one can easily add them up for method B markups or to desired level of commodity aggregation. The programs use M\(_1\) definition.

**SAS Program Files**
The files are stored in "/export/black-home2b/yao/home6/yao/hk8497". For years 1984-87, and 97, we do not have China data and there are only programs for method A (mkp**a.sas); for years 1988-96, there are programs for both methods A and C (mkp**ac.sas), where **=84-97. The first line of the program indicates the year this program was written for and the years it can apply to with minor change in input/output dataset names. Based on the similarities of data structure, three groups of programs are developed for methods A and C: programs for 1988-91, 1992-93 and 1994-96. Here we choose the 1988, 1992 and 1994 programs for an illustration (Appendix C).

The 1988 program has two parts: method A part and method C part. In method A part, it first constructs a quantity ratio indicator “qratio_t”, which measures the ratio of Hong Kong’s total quantity of a certain commodity re-exported from an origin country CTY_O to the rest of the world over the total import quantity imported from the same origin country. A dataset is created with this indicator, origin country codes and SITC code. Next, the dataset with the quantity ratio indicator is merged with import data and re-export data to form a dataset that has all the necessary information for markup calculation. In this merged dataset, we proceed to compute the markups (M) and create a dataset “mkp88a.ssd04” with the variables needed for further aggregation and subsequent method C calculation: commodity codes (SITC6), origin country codes (CTY_C), destination country codes (CTY_C), markup (PPCT_A), weight (WT_A) and re-export quantity (RQUANT_Y).

Also in part A, a statement “if qratio_t>2 then delete;” is included to eliminate records with measurement errors. One can choose to implement or ignore this statement and this is the only place in the SAS program one need to do something to delete the erroneous records. The method C markup calculations take method A markups as input data and one does not have to anything in method C part to delete the erroneous records.

The method C part of the program starts with preparing China’s export data. It first picks up all the records of China’s exports via Hong Kong (CTY_PS='110’) and converts the country codes assigned by China Customs to Hong Kong country codes. A country codes concordance file “cty_conc.ssd04” has been prepared for this conversion. The resulting dataset contains information on commodity, its unit, destination country, value and quantity of export via Hong Kong. Based on this dataset, it further computes the unit value of China’s export to a destination country via Hong Kong (UVHKCTY) and the unit value of China’s export via Hong Kong (UVHKROW), and the ratio of this two unit values (UVRATIO). Since Hong Kong’s 1988 data is classified at more disaggregate level (6-digit SITC) than China’s export data for the same year (6-digit SITC), it has to be aggregated to 5-digit SITC level. This work is done first by picking the records with China as the origin country from method A markup dataset “mkp88a.ssd04” and summing over the markup, weights and quantity to 5-digit SITC. By so doing, we assume that records with the same 5-digit SITC codes have the same unit codes. This may be true for most records in China data, but not for all. The resulting dataset has the following variables: SITC5, CTY_C, PPCT_A, WT_A and RQUANT_Y, which is compatible with the China data. When the two datasets are merged by SITC5, method X markup (PPCT_C) computed, which with other variables SITC5, CTY_C and WT_C, is
kept in dataset “mkp88c.ssd04”. Again, by merging China and Hong Kong data, we assume that the records at 5-digit SITC codes in the two datasets have the same unit, which is not true in general.

The 1994 program is pretty much similar to the 1988 program except that the former is classified with HS codes. It also has similar unit problems when the data is aggregated and merged.

The 1992 program, however, differs in method C part from the 1988 program, because the 1992 Hong Kong data is classified at 5-digit SITC while the 1992 China data is classified at 6-digit HS codes. Therefore, it is the China data that has to be first converted to 5-digit SITC codes, before being merged with the China data in method C part. A concordance between 5-digit SITC and 6-digit HS “conc.ssd04” is used to carry out this task. A problem with this conversion is that the China data comes with unit codes associated with the HS codes. Thus when converted to the SITC codes, there are cases that more than one unit codes are matched to one SITC codes. To avoid this problem, we keep only the unit that has the highest value within a HS category, then merge it with the Hong Kong data. But we still have the problem of different units for Hong Kong’s re-export quantity and China’s export quantity in the merged data.

The quantity unit problem is a problem since in method C, the ratio between China’s export quantity and Hong Kong’s re-export quantity is used to compute the unit value of Hong Kong’s import from China; thus the quantity information has to be properly aggregated and in the same unit when compared. But it does not matter that much to fundamentally change the markup calculation, especially the aggregated markup calculation, because the quantity ratio is only used as weights in the calculation of the import unit value $PM_i^*$ in method C. But as a matter of conceptual consistency, the unit problem would be better fixed for future implementations.

5. Conclusion

The novel approach to incorporate China export data into the markup calculations is made possible because China export data contains both the consignment country (purchasing country) and destination country codes. This gives more accurate measurement of the unit value for Chinese goods destined for a specific country via Hong Kong. Re-export trade is not a rare phenomenon in today’s international trade. Beside Hong Kong, Netherlands is also a re-export trade center in Europe. Re-export trade can distort the official trade data to such an extent that it renders the official data misleading for sensible research and policy discussions, and causes great confusion in understanding trade relations among countries, such as the US-China trade balance. To make the work easier on reconciling conflicting trade statistics caused by re-export trade, it is desirable that re-export ports compile accurate re-export data as Hong Kong does and countries much of whose exports are handled by a re-export center also have information on purchasing/destination countries included in their official data.

6. References