Data Processing for DOTS back up plan

Due to the lack of several countries in Com-trade database, we conducted the DOTS back up plan to calculate total trade value without breakdowns of commodities between each import country and partner country on the list of 187 countries.

We used IMF Direction of Trade Statistics (DOTS) database and Hong Kong re-export and re-import adjustment values for year 2015 and 2016 as data source for the plan.

We did data processing as following:

1. Removed 'TBG\_USD' level from ‘Indicator Code’ attribute and only keep 'TMG\_USD' and 'TXG\_USD' indicating import and export respectively. 'TBG\_USD' means value of trade balance, equal to the difference of import and export value.
2. Removed ‘status’ level from ‘Attribute’ column and removed ‘Attribute’ after it.
3. Transposed columns ‘X15’, ‘X16’, ‘X17’ and ‘X18’ into columns ‘year’ and ‘value’. And removed all the rows with null values generated because of transposing as well as all columns with names starting with ‘X’.
4. There’re some values in DOTS based on cost insurance and freight (CIF) while others based on free on board (FOB). Thus, we need to make all values on same basis for better comparing and calculating. We deflated values based on CIF by 1.06 considering that IMF indicates 6% as CIF factor.
5. Split dataset by trade directions into import and export datasets.
6. Merged import by export dataset on year, report country and partner country. All records containing null values generated here are the orphaned records. Repeated previous merging step in opposite direction and get a list of lost records.
7. Removing orphaned records from import dataset and lost from export, we got matched trade records for both import and export.
8. Converted UN code in Hong Kong data into IMF code, origin country into export country and consign country into import country. Transferred HK dollars to USD by IMF yearly average exchange rates. Then concatenated two Hong Kong datasets for 2015 and 2016.
9. Merged Hong Kong data into matched import and export datasets respectively by year, report country and partner country. Kept all null values in ‘value’ attribute as 0 because we didn’t merge Hong Kong adjustment data for all years to DOTS, but only for year 2015 and 2016. And removed re-exports and re-imports from DOTS value to get adjusted trade values following two formulas below:

* Adjusted import value = Original import value from report country – Original export value from partner country – Hong Kong re-import value
* Adjusted export value = Original import value from partner country – Original export value from report country - Hong Kong re-export value

1. Merged adjusted and matched import and export datasets together as final database by year, report country, partner country as well as indicator code.
2. Created a new column to indicate positive values as over mis-invoicing and negative as under mis- invoicing. Then converted all values into positive numbers.

Based on steps above, we got a DOTS database with matched trade records from year 2015 to year 2018 and it has been adjusted by Hong Kong re-export and re-import for year 2015 and 2016. We also got two lists of orphaned and lost trade records.