Complete TF CPC

Author: Alex Matrunich, Marco Garieri, Bo Werth, Christian Mongeau

Description:

The trade module is divided in two submodules: **complete_tf_cpc** and **total_trade_CPC**. Each module is year specific. This means that, at the time being, the trade module run indipendently for each year. In order to run the **tt total_trade_CPC**, the output of **complete_tf_cpc** is needed.

Flow chart:

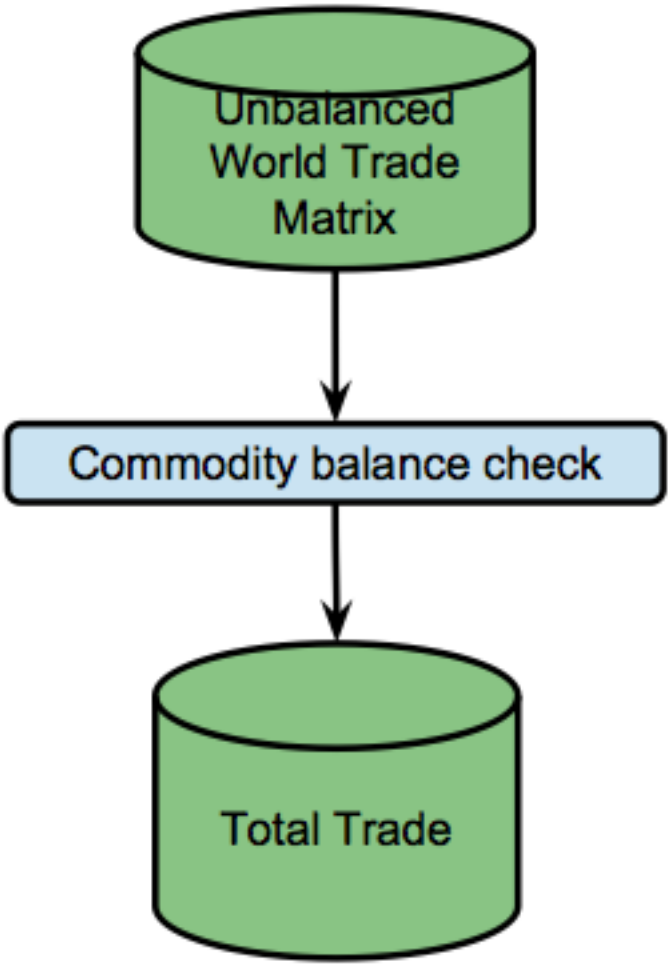

Figure 1: Aggregate complete_tf to total_trade

Input Data

Supplementary Datasets:

1. `hsfclmap2`: Mmapping between HS and FCL codes extracted from MDB files used to archive information existing in the previous trade system (Shark, Jellyfish).

2. `adjustments`: Adjustment notes containing manually added conversion factors to obtain quantities from traded values

3. `unsdpartnersblocks`: UNSD Tariffline reporter and partner dimensions use different list of geographic are codes. The partner dimesion is more detailed than the reporter dimension. Since we can not split trade flows of the reporter dimension, trade flows of the corresponding partner dimensions have to be assigned the reporter dimension's geographic area code. For example, the code 842 is used for the United States includes Virgin Islands and Puerto Rico and thus the reported trade flows of those territories. Analogous steps are taken for France, Italy, Norway, Switzerland and US Minor Outlying Islands.

4. `fclunits`: For UNSD Tariffline units of measurement are converted to meet FAO standards. According to FAO standard, all weights are reported in metric tonnes, animals in heads or 1000 heads and for certain commodities, only the value is provided.

5. `comtradeunits`

6. `EURconversionUSD`: Annual EUR/USD currency exchange rates table from SWS

Extract UNSD Tariffline Data

1. Chapters: The module downloads only records of commodities of interest for Tariffline Data. The HS chapters are the following: 01, 02, 03, 04, 05, 06, 07, 08, 09, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 33, 35, 38, 40, 41, 42, 43, 50, 51, 52, 53. In the future, if other commotidy are of interest for the division, it is important to include additional chapter in the first step of the downloading. For Eurostat Data no filtering is applied.

2. Remove duplicate values for which quantity & value & weight exist (in the process, removing redundant columns). Note: missing quantity|weight or value will be handled below by imputation

3. The tariffline data from UNSD contains multiple rows with identical combination of reporter / partner / commodity / flow / year / qunit. Those are separate registered transactions and the rows containinig non-missing values and quantities are summed.

4. Remove non-numeric comm (hs) code; comm (hs) code has to be digit. This probably should be part of the faoswsEnsure

Extract Eurostat Combined Nomenclature Data

1. Remove reporters with area codes that are not included in MDB commodity mapping area list
2. Convert HS to FCL
3. Remove unmapped FCL codes
4. Join *fclunits*
5. `NA` *fclunits* set to `mt`
6. Specific ES conversions: some FCL codes are reported in Eurostat with different supplementary units than those reported in FAOSTAT

Harmonize UNSD Tariffline Data

1. Geographic Area: UNSD Tariffline data reports area code with Tariffline M49 standard (which are different for official M49). The area code is converted in FAO country code using a specific convertion table provided by Team ENV. Area codes not mapping to any FAO country code or mapping to code 252 (which correpond not defined area) are separately saved and removed from further analyses.

2. Commodity Codes: Commodity codes are reported in HS codes (Harmonized Commodity Description and Coding Systpem). The codes are converted in FCL (FAO Commodity List) codes. This step is performed using table incorporated in the SWS. In this step, all the mapping between HS and FCL

code is stored. If a country is not included in the package of the mapping for that specific year, all the records for the reporting country are removed. All records without an FCL mapping are filtered out and saved in specific variables.

3. Aggregate UNSD Tariffline Data to FCL: here we select column `qtyfcl` which contains weight in tons (requested by FAO).

Combine Trade Data Sources

1. The adjustment notes developed for national data received from countries are not applied to HS data any more (see instructions 2016-08-10). Data harvested from UNSD are standardised and therefore many (if not most) of the quantity adjustment notes (those with no year) need not be applied. The "notes" refer to the "raw" non-standardised files that we used to regularly receive from UNSD and/or the countries. Furthermore, some data differences will also arise due to more recent data revisions in these latest files that have been harvested.
2. Convert currency of monetary values from EUR to USD using the `EURconversionUSD` table (see above).
3. Combine UNSD Tariffline and Eurostat Combined Nomenclature data sources to single data set.

- TL: assign `weight` to `qty`
- ES: assign `weight` to `qty` if `fclunit` is equal to `mt`, else keep `qty`

Outlier Detection and Imputation

1. Unit values are calculated for each observation at the HS level as ratio of monetary value over weight `value / qty`.
2. Median unit-values are calculated across the partner dimension by year, reporter, flow and HS. This can be problematic if only few records exist for the a specific combination of dimensions.
3. Observations are classified as outliers if the calculated unit value for a some partner country is below or above the median unit value. More specifically, the measure defined as median inter-quartile-range (IQR) multiplied by the outlier coefficient (default value: 1.5) is used to categorize outlier observations.
4. Impute missing quantities and quantities categorized as outliers by dividing the reported monetary value with the calculated median unit value.
5. Assign `flagTrade` to observations with imputed quantities. These flags are also assigned to monetary values. This may need to be revised (monetary values are not supposed to be modified).
6. Aggregate by FCL over HS dimension: reduce from around 15000 commodity codes to around 800 commodity codes.
7. Map FCL codes to CPC, remove observations that have not been mapped to CPC.

Mirror Trade Estimation

1. Obtain list of non-reporting countries as difference between the list of reporter countries and the list of partner countries.
2. Swap the reporter and partner dimensions: the value previously appearing as reporter country code becomes the partner country code (and vice versa).
3. Invert the flow direction: an import becomes an export (and vice versa).
4. Calculate monetary mirror value by adding a 12% mark-up on imports to account for the difference between CIF and FOB prices.

5. In this step, no new flags are assigned explicitly. Imputation flags created before are copied to new records.
6. Assign SWS ObservationStatus flag `I` and flagMethod `e` to records with with `flagTrade` unless the FCL unit is categorized as `$ value only`.

Output for SWS

1. Filter observations with FCL code `1181` (bees).
2. Filter observations with missing CPC codes.
3. Rename dimensions to comply with SWS standard
4. Transform dataset seperating monetary values and quantities in different rows.
5. Convert values and quantities to corresponding SWS element codes. For example, a qunatity import with unit "mt" is assigned `5610`.