Appendix: `complete_tf_cpc` module

Marco Garieri
Alexander Matrunich
Christian A. Mongeau Ospina
Bo Werth
Food and Agriculture Organization of the United Nations

5 September 2017

This document gives a faithful step-by-step sequence of the operations performed in the `complete_tf_cpc` module. For a narrative version of the module's approach, please see its main document.

Parameters

- `year`: year for processing.
- `out_coef`: coefficient for outlier detection, i.e., the `k` parameter in the *Outlier Detection and Imputation* section.

```
## INFO [2017-09-05 15:27:00] HS chapters to be selected:
##
## [1] "'01', '02', '03', '04', '05', '06', '07', '08', '09', '10', '11', '12', '13', '14', '15', '16',
```

- `hs_chapters`: can not be set by the user as it is provided by Team B/C and harcoded). 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', '43', '50', '51', '52', '53'

Download raw data and basic operations

1. Download Eurostat data (ES)
2. Download Tariff line data (TL)
3. Keep only `stat_regime` = 4 for ES
4. Remove European-aggregated data (i.e., totals) from ES
5. Use standard (common) variable names (e.g., `declarant` becomes `reporter`) in ES and TL.
6. Remove non numeric reporters / partners / hs codes from ES and TL.
7. Use standard (common) variable types in ES and TL.
8. Filter HS codes of interest, i.e., codes that don't participate in further processing. Such solution drops, e.g., all HS codes shorter than 6 digits.
9. Convert ES geonomenclature country/area codes to FAO codes.
10. TL M49 codes (which are different from official M49) are converted in FAO country codes using a specific convertion table provided by Team ENV. See below for the description of the `unsdpartnersblocks` table.
11. Remove invalid repoters (i.e., keep contries/areas that existed in the year considered).
12. Remove ES repoters from TL.

Loading of help datasets

- `hsfclmap3`: Mapping between HS and FCL codes extracted from MDB files used to archive information existing in the previous trade system (Shark/Jellyfish). This mapping is provided by a separate package: https://github.com/SWS-Methodology/hsfclmap Unmapped codes in this table are supplemented by newly created links stored in the `hsfclmap4` table.
- `hsfclmap4`: New mapping between HS and FCL codes (extends `hsfclmap3`)
- `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.
- `fclunits`: For UNSD Tariffline units of measurement are converted to meet FAO standards. According to FAO standard, all weights are reported in tonnes, animals in heads or 1000 heads and for certain commodities, only the value is provided.
- `comtradeunits`: Translation of the `qunit` variable (supplementary quantity units) in Tariffline data into intelligible unit of measurement, which correspond to bthe standards of quantity recommended by the *World Customs Organization* (WCO) (e.g., `qunit`=8 correspond to *kg*). See: http://unstats.un.org/unsd/tradekb/Knowledgebase/UN-Comtrade-Reference-Tables
- `EURconversionUSD`: Annual EUR/USD currency exchange rates table from SWS.

Generate HSFCLMAP6 map

1. Add variables that will contain flags. (Note: flags are set in various steps in the code. Please, refer to the "flag_management.docx" document.)
2. Generate Observation Status "X" flag and Method "h" flag
3. Remove code 252
4. Remove reporters with area codes that are not included in MDB commodity mapping area list.
5. Map HS to FCL, ES.
6. Remove unmapped FCL codes.
7. Add FCL units.
8. Specific ES conversions: some FCL codes are reported in Eurostat with different supplementary units than those reported in FAOSTAT, thus a conversion is done.
9. Do mathematical conversions on specific `qunits` (6, 9, and 11 become 5)
10. Identical combinations of reporter / partner / commodity / flow / year / qunit are aggregated.
11. Add variables that will contain flags.
12. Generate Observation Status "X" flag and Metdoh "h" flag.
13. Area codes not mapping to any FAO country code are removed.
14. Re-imports become imports and re-exports become exports.
15. Map HS to FCL, TL.
16. Add FCL units.

17. General TL conversions: some FCL codes are reported in Tariffline with different units than those reported in FAOSTAT, thus a conversion is done.
18. Specific TL conversions: some commodities need a specific conversion.
19. If the `weight` variable is available and the final unit of measurement is tonnes then `weight` is used as `quantity`
20. Aggregate UNSD Tariffline Data to FCL.

Combine Trade Data Sources

1. Application of "adjustment notes" to both ES and TL data.
2. Convert currency of monetary values from EUR to USD using the `EURconversionUSD` table.
3. Assign 'weight' flags to 'qty' flags in TL XXX.
4. Assign `qty` flags to `weight` flags in ES but only when `fclunit` is different from "mt".
5. 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 "mt", else keep `qty`

Imputation

1. Unit values are calculated for each observation at the HS level as ratio of monetary value over quantity `value / qty`.
2. Imputation of missing quantities and quantities categorized as outliers by applying the method presented in the *Missing Quantities Imputation* section. The `flagTrade` variable is given a value of 1 if an imputation was performed.

Separate flags.

1. Aggregate values and quantities by FCL codes.
2. Set flags for aggregated values/quantities.
3. Map FCL codes to CPC.
4. Map FAO area codes to M49.
5. Countries with FAOSTAT code 252 are converted to M49 896 ("Other nei").

Mirror Trade Estimation

1. Create a table with the list of reporters and partners combined as areas and count the number of flows that the areas declare as reporting countries. The partners that never show up as reporters or the reporters that do not report a flow will have a number of flows equal to zero and will be mirrored.
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 (removing) a 12% mark-up on imports (exports) to account for the difference between CIF and FOB prices.
5. Set flags to mirrored flows.

Flag aggregation

Flags are aggregated as explained in the *Flags* section in the main documentation.

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, e.g., `geographicAreaM49Reporter`
4. Calculate unit value (US$ per quantity unit) at CPC level if the quantity is larger than zero
5. Transform dataset separating monetary values, quantities and unit values in different rows.
6. Convert monetary values, quantities and unit values to corresponding SWS element codes. For example, a quantity import measured in metric tons is assigned `5610`.