Agricultural trade data processing

July 9, 2015

Last update: 2015-08-10 10:12:14

Raw trade data

FAO receives data on trade flows from United Nations Statistical Division. The division runs Commodity Trade Statistics Database UN Comtrade "It stores standardised official annual trade statistics reported by countries and reflecting international merchandise flows detailed by commodity and partner country with coverage reaching up to 99 percent of world merchandise trade[1]".

One can freely download this standardised statistics from the open data base. Statistical Division of FAO gets unstandardised data.

Table 1: Random sample of import trade flows of 2011 year, reported by the US

year	reporter	partner	hs	flow	weight	qty	qunit	value
2011	842	246	4015900050	1	NA	2	11	399
2011	842	360	6309000010	1	1242	1242	8	462
2011	842	703	8528727210	1	NA	2	5	5150
2011	842	442	8514908000	1	NA	NA	1	3387
2011	842	788	6211420081	1	1162	366	11	87308
2011	842	699	2931009021	1	309000	309000	8	6914048
2011	842	276	3823130040	1	28810	28810	8	90332
2011	842	616	8531909000	1	NA	NA	1	78492
2011	842	372	8007001050	1	NA	NA	1	18784
2011	842	458	4203293010	1	NA	3960	6	7597

This is an example of unstandardised data on trade inflows in 2011, reported by the United States. Reporters and trade partners are represented with three-digit numerical codes used by the Statistics Division of the United Nations. Trade commodities are classified with extended Harmonized Commodity Description and Coding System (HS)[2] maintained by the World Customs Organization[3].

Weight is measured in kilograms and value in US dollars. Quantity (qty column) is an optional alternative for weight. It could be measured in different units (qunit column). See full list of possible units and their descriptions in Annex I of Quantity and Weight Data in UN Comtrade[4].

Country-specific HS commodity codes

Harmonized system classification is declared by WCO up to 6 digits. A country may extend HS to more detailed level to better respond to local circumstances. Let's compare differencies in codes under subheading

[1] http://unstats.un.org/unsd/comtrade_announcement.htm Comtrade Announcement

[2] http://www.wcoomd.org/en/topics/nomenclature/overview/what-is-the-harmonized-system.aspx What is the Harmonized System (HS)?

[3] http://www.wcoomd.org/en.aspx World Customs Organization

[4] http://unstats.un.org/unsd/tradekb/Knowledgebase/Quantity-and-Weight-Data-in-UN-Comtrade Quantity and Weight Data in UN Comtrade

0202 Meat of bovine animals, frozen between the US and Brazil[5].

Table 2: Extension of HS codes by the US

hs	Description
02	CHAPTER 2 MEAT AND EDIBLE MEAT OFFAL
0202	Meat of bovine animals, frozen:
0202.10	- Carcases and half-carcases:
0202.10.05	- - Described in general note 15 of the tariff schedule . . . :
0202.10.05.10	- - - Veal
0202.10.05.90	- - - Other
0202.10.10	- - Described in additional U.S. note 3 to this chapter . . . :
0202.10.10.10	- - - Veal
0202.10.10.90	- - - Other
0202.10.50	- - Other:
0202.10.50.10	- - - Veal
0202.10.50.90	- - - Other
0202.20	- Other cuts with bone in:
	- - Described in general note 15 of the tariff schedule . . . :
	- - - Processed:
0202.20.02	- - - - High-quality beef cuts
0202.20.04	- - - - Other
0202.20.06	- - - Other
	- - Described in additional U.S. note 3 to this chapter . . . :
	- - - Processed:
0202.20.10	- - - - High-quality beef cuts
0202.20.30	- - - - Other
0202.20.50	- - - Other
0202.20.80	- - Other
0202.30	- Boneless:
	- - Described in general note 15 of the tariff schedule . . . :
	- - - Processed:
0202.30.02	- - - - High-quality beef cuts
0202.30.04	- - - - Other
0202.30.06	- - - Other
	- - Described in additional U.S. note 3 to this chapter . . . :
	- - - Processed:
0202.30.10	- - - - High-quality beef cuts
0202.30.30	- - - - Other
0202.30.50	- - - Other
0202.30.80	- - Other

Table 3: Extension of HS codes by Brazil

hs	Description
02	CHAPTER 2 MEAT AND EDIBLE MEAT OFFAL
0202	Meat of bovine animals, frozen:
0202.10	- Carcases and half-carcases
0202.20	- Other cuts with bone in:
0202.20.10	- - Forequarters

[5] http://madb.europa.eu Descriptions of country-specific HS-codes are provided by Market Access Database and copyrighted by Mendel Verlag, Germany.

hs	Description
0202.20.20	- - Hindquarters
0202.20.90	- - Other
0202.30	- Boneless

The set of HS-codes from the US is wider, than Brazilian one. For boneless meat Brazil doesn't extend standard code 0202.30, when the US use here seven additional codes.

Country codes

Codes of reporters

Area codes of reporters are standardized by the Statistical Department. The SD follows *in general* the United Nations Standard Country or Area Codes for Statistical Use[6]. The code scheme used by the SD[7] is slightly modified from the official one[8]. For example the official scheme offers code 840 for the US, when the modified version uses 842.

Codes of partners

Partners' codes in Tariffline data are not standardised and presented as they were reported by countries. Reporters can use as standard version of codes, so the version of the Statistical Department. For example, in Tariffline data there are 27 country codes which are not presented in official scheme and 40 codes not covered by the modified version.[9]

Initial validation of trade data

At prevalidation step we are to make a decision should we accept data from a specific country for the further processing or not. A country could provide data of good quality for one part of commodities and inadequate level of quality for another part. We want to estimate quality differences between commodities of a country.

Quality of data is estimated by following indicators:

- Share per cent of missing quantities
- Share per cent of unit value outliers

Self-trade

There are cases when a country reports itself as a partner to exports or imports. Such situations can occur due to mistakes or when an entrepôt exists.

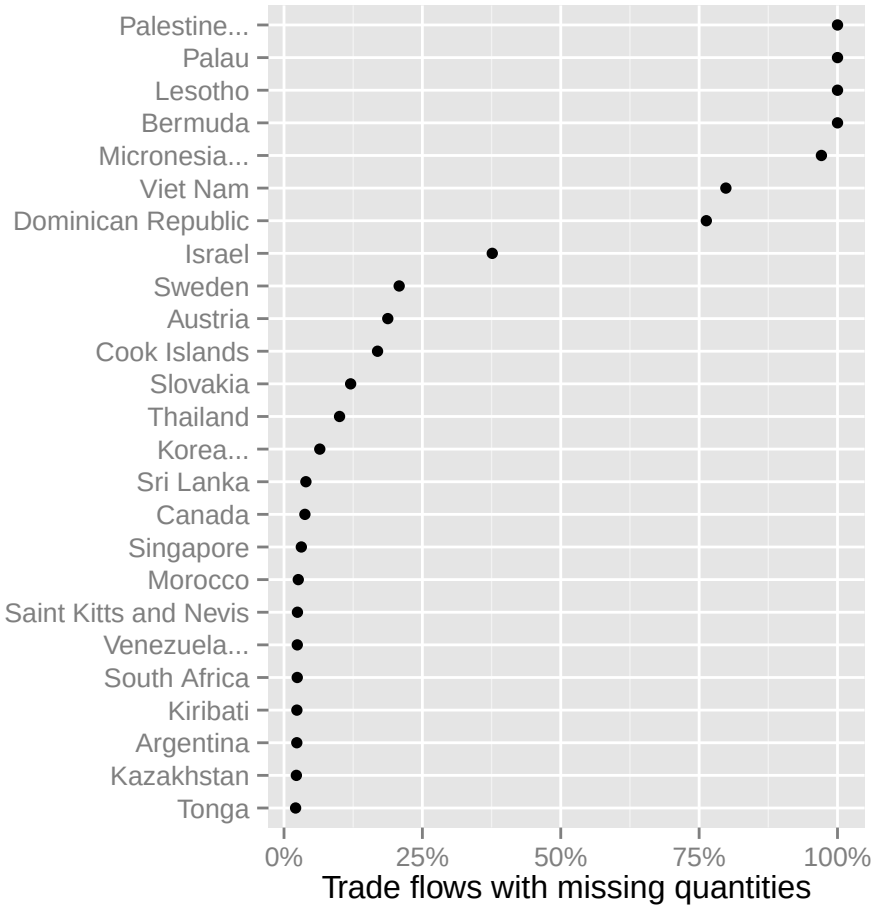
[6] http://comtrade.un.org/pb/ The United Nations Statistics Division (2015). The 2014 International Trade Statistics Yearbook, Volume I - Trade by Country, xix.
[7] http://comtrade.un.org/data/doc/api/ The UN Comtrade data extraction API
[8] http://unstats.un.org/unsd/methods/m49/m49alpha.htm Countries or areas, codes and abbreviations
[9] http://rpubs.com/malexan/m49 Matrunich A. (2015). M49 country codes in Tariffline

Table 4: Self-trade of commodities from 2nd, 10th and 15th HS chapters in 2011

Reporter	Flow	Total
France	Import	163
Canada	Import	56
Portugal	Import	50
Slovakia	Import	50
New Zealand	Import	31
South Africa	Import	27
United Kingdom	Import	17
Slovenia	Import	16
Estonia	Import	12
Thailand	Import	12
China	Import	7
Greenland	Import	3
Ethiopia	Import	2
Papua New Guinea	Import	2
Saint Kitts and Nevis	Import	2
Indonesia	Import	1
Malaysia	Import	1
Palau	Import	1

Missing quantity

We identify which reporters provide data of insufficient quality. Firstly for every reporter proportion of trade flows with missing quantity is calculated.

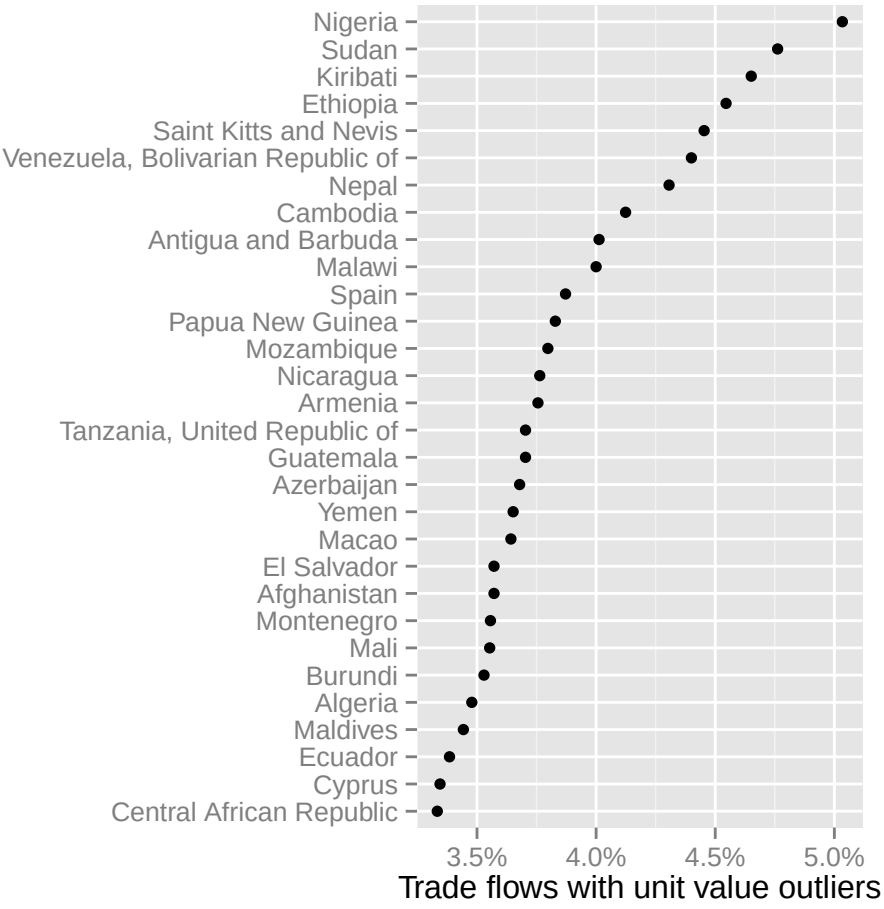


Detection of outliers

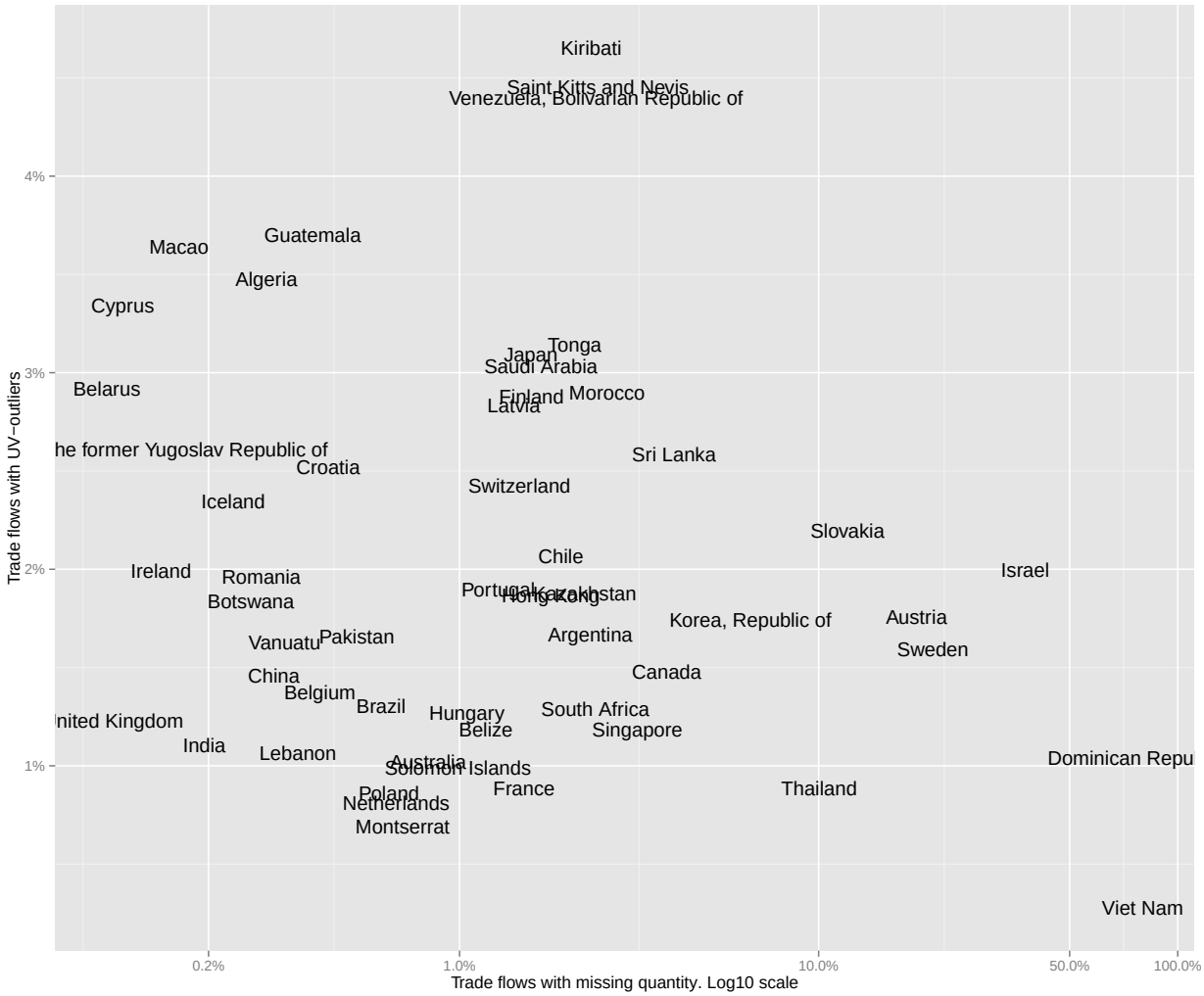
We define outliers as observations located outside the range:

$$[Q_1 - k(Q_3 - Q_1), Q_3 + k(Q_3 - Q_1)]$$

where Q_1 and Q_3 are the lower and upper quartiles respectively, and k is a non negative constant. In this paper we use k equals 1.5.


Nigeria
Sudan
Kiribati
Ethiopia
Saint Kitts and Nevis
Venezuela, Bolivarian Republic of
Nepal
Cambodia
Antigua and Barbuda
Malawi
Spain
Papua New Guinea
Mozambique
Nicaragua
Armenia
Tanzania, United Republic of
Guatemala
Azerbaijan
Yemen
Macao
El Salvador
Afghanistan
Montenegro
Mali
Burundi
Algeria
Maldives
Ecuador
Cyprus
Central African Republic
3.5%
4.0%
4.5%
5.0%
Trade flows with unit value outliers

Missing quantites and outliers combined



Imputing of missing quantities and replacement of outliers

In data reported by USA for 2011 year in HS chapters 2, 10 and 15 there are 0 trade flows with missing quantity and 33 trade flows with UV-outliers.

Table 5: Example trade flows with outlied unit values

Reporter	Partner	Flow	Commodity	Weight	Value	UV	UV_me
United States	Japan	Import	1008200000	327	2479	7.58	1.88
United States	Japan	Import	1516100000	1090	93260	85.56	27.54
United States	United Kingdom	Import	1521100040	84	6181	73.58	6.89
United States	Japan	Import	1001902096	61	46362	760.03	1.47
United States	Canada	Import	1509102000	58	2386	41.14	5.93
United States	Japan	Import	1006204040	2520	9679	3.84	0.89
United States	South Africa	Import	1006204080	162	4752	29.33	1.22
United States	France	Import	1510004000	279	3517	12.61	2.45

Reporter	Partner	Flow	Commodity	Weight	Value	UV	UV_me
United States	United Kingdom	Import	1514999010	2290	24200	10.57	2.67
United States	Netherlands	Import	1504204000	160571	4932146	30.72	5.83

Imputing using reporter median unit values

Now we correct weight of this outlied trade flows with help of median reporter unit value for a given commodity.

Table 6: Example trade flows with corrected weight

Reporter	Partner	Flow	Commodity	Weight	Weight_corr	Weight_diff
United States	Japan	Import	1008200000	327	1319	-992
United States	Japan	Import	1516100000	1090	3386	-2296
United States	United Kingdom	Import	1521100040	84	897	-813
United States	Japan	Import	1001902096	61	31539	-31478
United States	Canada	Import	1509102000	58	402	-344
United States	Japan	Import	1006204040	2520	10875	-8355
United States	South Africa	Import	1006204080	162	3895	-3733
United States	France	Import	1510004000	279	1436	-1157
United States	United Kingdom	Import	1514999010	2290	9064	-6774
United States	Netherlands	Import	1504204000	160571	845994	-685423

Imputing using data from trade partner (mirroring)