Standard Unit Values

1. Introduction

- A **'standard unit value'** (SUV) is defined for each commodity at the 6-digit level of aggregation, by year, type of trade flow (imports/exports), and different quantity units.

- The SUV serves two main **objectives**:
  1. It is used to **estimate volume** of trade when only monetary values are available
  2. It also provides a **benchmark** against which the quality of new value/volume data pairs can be assessed.

- A **sample** of unit values for each commodity, flow, and year is available in **COMTRADE** when dividing total values by their respective quantities. Based on that sample, a Standard Unit Value (SUV) can be calculated for each commodity/flow/year, using the median unit value of value/quantity ratios.

- The methodology to calculate Standard Unit Values can be applied for several commodity classifications. At the moment, work has been completed for HS0, HS1, and HS0 classifications, and there is work in progress for the different revisions of the SITC classification.

- Section 2 of this report summarizes some **features of the unit value data** using descriptive statistics. It provides alternative **measures of location, dispersion, and skewness** for the sample distribution of unit values, and illustrates these findings with some specific examples.

- The **main conclusions** from the descriptive analysis of Section 2 are:

  1. Unit value data for most commodities exhibit **high degree of variability**

  2. The distribution of unit values is usually **asymmetric** around its mean (skewness is usually positive

  3. The data is affected by the presence of **outliers.**

  4. A **log-transformation** of the unit value data significantly reduces asymmetry, and therefore is more appropriate **to construct confidence intervals and rejection thresholds for outliers**.

- Section 3 sets up the criteria used to determine whether the available sample of unit values of a specific commodity/flow/year can be relied upon to determine a Standard Unit Value. Such criteria impose maximum acceptable limits on the asymmetry, spread and/or multimodality of the sample distribution. It also contains a list of the SQL scripts used to create Standard Unit Value tables for different commodity classifications.

2. Descriptive Statistics: Main Features of Unit Value Data

- [The descriptive statistics discussed in this Section are available for of each commodity/flow/year from 2000 to 2004 in the excel file 'DescriptiveStatisticsAll.xls']

2.1. Assessment of variability

- A first measure of variability in unit value data for each commodity/flow/year sample is their relative standard deviation, RSD, which is defined as the ratio of the standard deviation (s) divided by the arithmetic mean ($\bar{x}$)

- An analogous non-parametric measure of variability is the relative interquartile range, which is defined as

$$\mathrm{RIQ} = \frac{(Q_3 - Q_1)}{M},$$

where M represents the median and Q_1 and Q_3 the 25th and 75th percentiles of the unit value sample, respectively.

- For a majority of commodities, the unit values calculated on the basis of value and quantity data exhibit a high degree of variability, as measured by the relative interquartile range (see Figure 1).

- In particular, 65% of 51,945 commodity-specific unit value samples available for imports and exports in the period 2000-2004 (using only the recommended quantity units of measurement) have a relative interquartile range greater than one. This variability shall be taken into account when assessing the reliability of commodity-specific Standard Unit Values for volume estimation and quality-checks purposes.
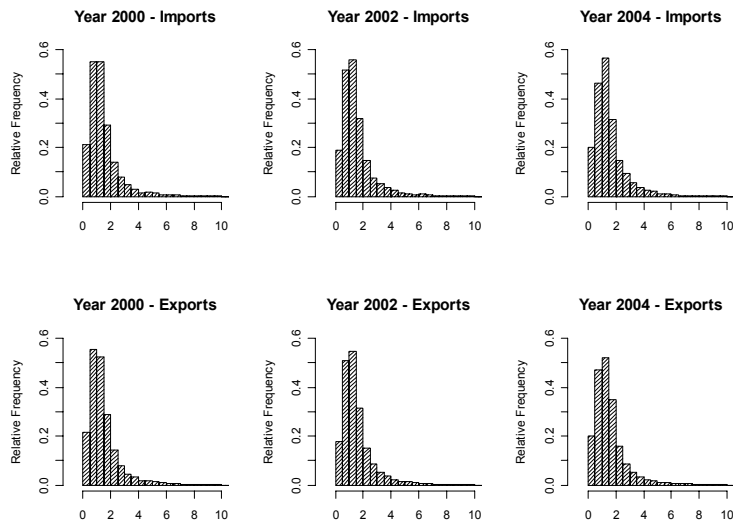
Year 2000 - Imports	Year 2002 - Imports	Year 2004 - Imports

Year 2000 - Exports	Year 2002 - Exports	Year 2004 - Exports

Figure 1. Distribution of the relative interquartile range of unit values among commodities

2.2. Assessment of asymmetry

- A non-parametric measure of skewness (or asymmetry) in the distribution of each unit value sample is provided by the Bowley skewness coefficient, which is defined as

$$\text{B} = \frac{(Q_3 - M) - (M - Q_1)}{Q_3 - Q_1} = \frac{(Q_3 - 2M + Q_1)}{Q_3 - Q_1},$$

Its value is bounded between -1 and +1, and it is equal to zero if the median is located exactly in the middle of the interquartile range.

- Examination of the samples (see Figure 2) reveals that the distribution of unit value data is typically skewed to the right (i.e., B > 0). More specifically, 92% of the commodity/flow/year samples of unit values have a positive Bowley coefficient, and in about 50% of the samples this coefficient is greater than 0.32.
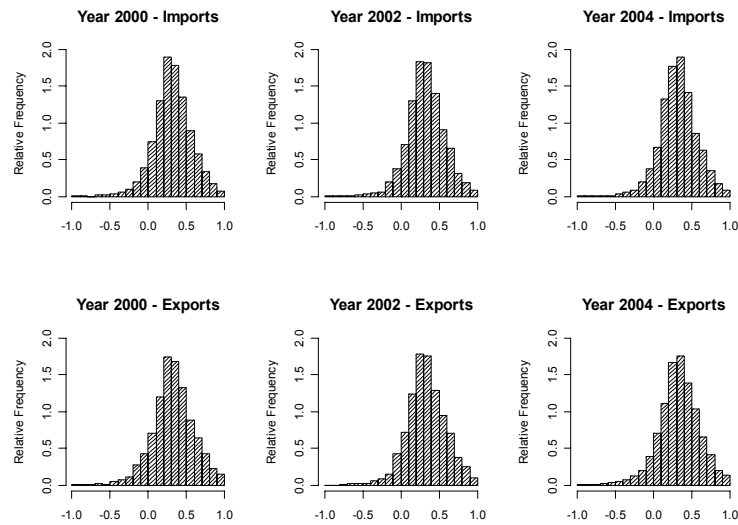
Figure 2. Distribution of the Bowles skewness of unit values among different commodities

- After applying a **logarithmic transformation** to the unit value data in each commodity/flow/year sample, their skew is typically near zero, as is shown in Figure 3. Moreover, approximately 50% of the transformed unit value samples have a Boewley coefficient of skewness that is bounded between -0.10 and 0.17, indicating that the logarithmic transformation is successful in restoring symmetry.
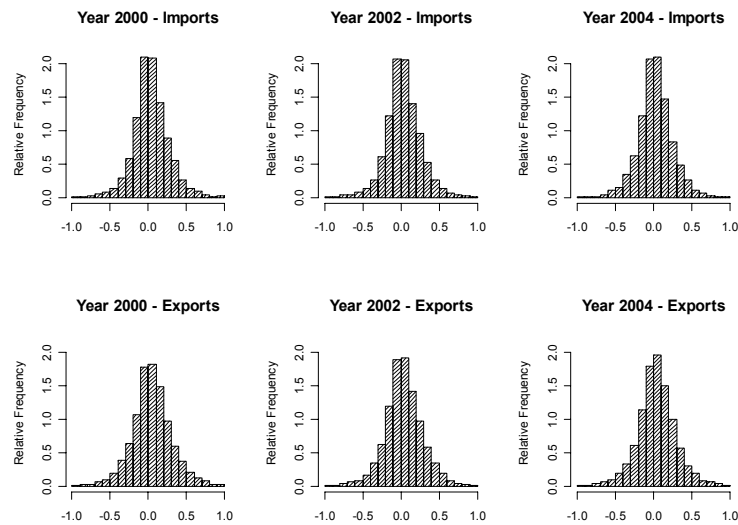

Figure 3. Distribution of the Bowles skewness of unit values among different commodities, after applying logarithmic transformation

2.3. Identification of outliers

- Data points that seem to be inconsistent with the general characteristics of the sample are called **outliers**. These are values "that lie far from the middle of the distribution in either direction."

- Outliers may arise for several reasons:
  1. **Errors** in data entry or processing.
  2. **Atypical circumstances** in the data generating process
  3. **Intrinsic variability** of the data generating process.

- Methods of outlier detection are useful for both conducting **data quality checks** and understanding the **reliability** and **intrinsic characteristics** of the data generating process.

- The method for outlier detection adopted in this report is based on the idea that most values are expected in the **interquartile range**, which is the interval between Q_1 and Q_3.

- On the log-transformed sample, the left and right **thresholds for anomalous values** are determined by adding to or subtracting from Q_1 or Q_3, respectively, a symmetric step equal to one and a half times the interquartile range.

- Using this criterion, about 4.7% of the observations in the unit value samples were diagnosed as outliers and disregarded from further calculations to obtain Standard Unit Values.

2.4. Assessing multimodality

- Determining a single Standard Unit Value from for the all transactions classified under a single commodity/flow/year is problematic if the data sample comes from various **heterogeneous** subpopulations.

- This form of heterogeneity is frequently reflected in the **presence of multiple modes** in the sample of individual unit values used to calculate a Standard Unit Value. Ideally, Standard Unit Values should be calculated from uni-modal samples.

- To assess the degree of multimodality in the samples of unit values available for each commodity/flow/year, the following **multimodality index** based on the histogram of the log-transformed data is proposed[1]:

---

[1] In defining the multimodality index, the histogram of the log-transformed data is constructed by assigning each data point to one of **ten equally-spaced cells** on the log-transformed scale.

$$\text{Multimodality index} = \frac{\left(m_1 + \cdots + m_k\right)^2}{m_1^2 + \cdots + m_k^2},$$

where k is the number of modes in the sample histogram and m_j is the mass weight attached to its jth mode (i.e., the number of data points falling in jth mode's cell, divided by the total number of individual unit values used to construct the histogram). If there is only one mode (i.e., if $k = 1$), the multimodality index takes the value of one; if there are two equally relevant modes (i.e., if $k = 2$, with $m_1 = m_2$), the index is equal to two; etc.

2.5. Some specific examples

- The following examples refer to export unit values in 2004 of several commodities. They provide an overview of the main features typically encountered in unit value data.

- In each table, the outlier detection criteria discussed above is applied to the sample of unit values for the corresponding commodity. Measures of location, spread, skewness, and multimodality are also presented, both before and after removing outliers.

- The left plot under each table contains the **histograms of the unit value data before removing outliers** (in logarithmic scales), as well as a **box-plot** indicating:
  1. The location of the **interquartile range** (the length of the "box")
  2. The location of the **median** (the bold vertical line dividing the box in two parts)
  3. The location of the **acceptance thresholds** that are used to detect outliers (represented by the extremes of the "whiskers").

- The plot to the right shows the **histograms of the unit value data after removing outliers** (in logarithmic scale).
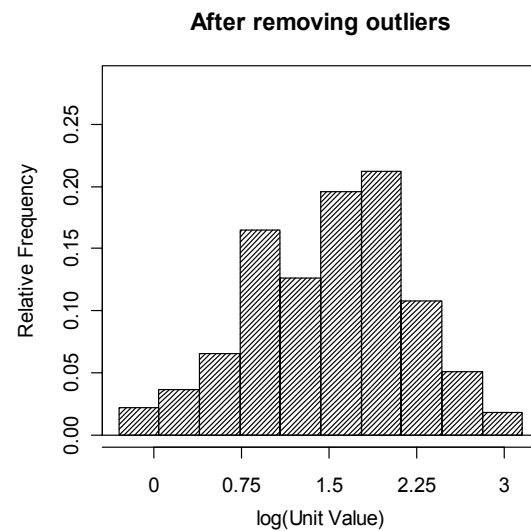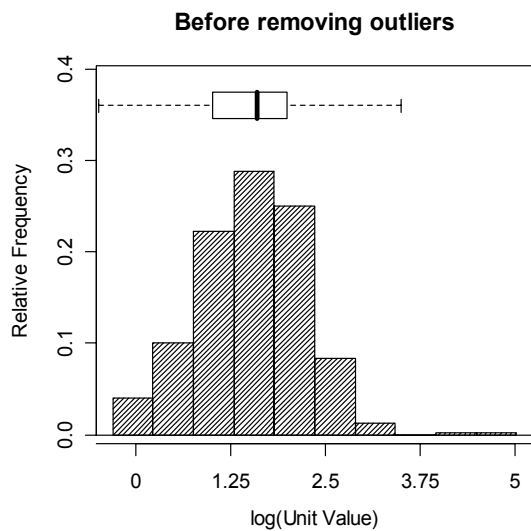
HS2 010110 - Live horses/asses/mules/hinnies: pure-bred breeding animals (Quantity unit: 5)

Number of observations:	**116**	
Total quantity:	245,279	
Total value:	704,651,457	
Number of left outliers:	**1**	
Total quantity:	218,927	
Total value:	261,771	
Number of right outliers:	**0**	
Total quantity:		
Total value:		
Detection of outliers		
Left threshold:	100.23	
Right threshold:	1,070,785.29	
Descriptive statistics	Before removing outliers	After removing outliers
Min:	1.20	135.02
Q1:	3,249.17	3267.41
Median:	11,333.80	11418.23
Q3:	33,032.71	33138.91
Max:	1,057,588.00	1,057,588.00
Arithmetic mean:	43,607.90	43,987.09
Geometric mean:	10,086.37	10,911.26
Bowley measure of skewness		
Original data:	0.46	0.45
Log-transformed data:	-0.08	-0.08
Multimodality index:	1.05	1.23

Before removing outliers

After removing outliers
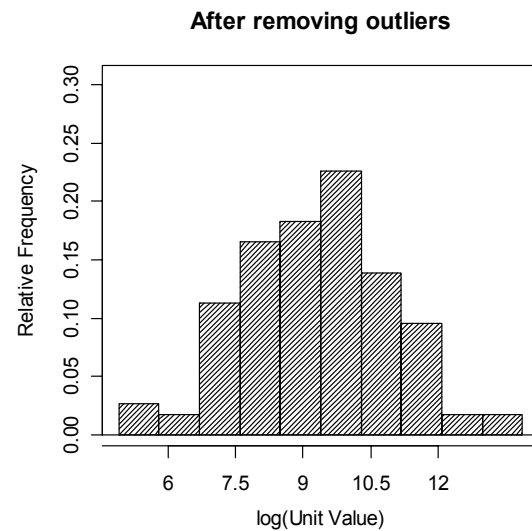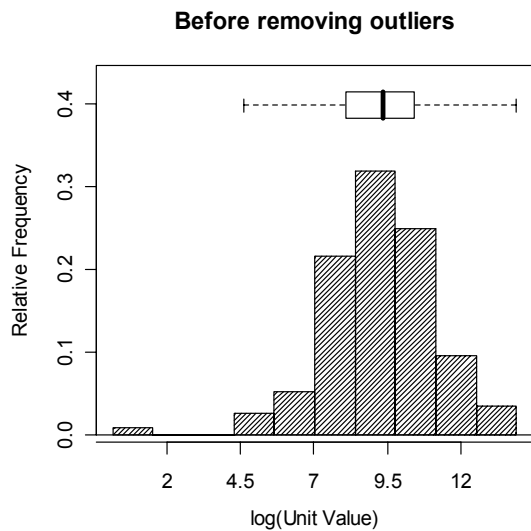
HS2 020130 - Meat of bovine animals, fresh/chilled, boneless (Quantity unit: 8)

Number of observations:	**549**	
Total quantity:	1,654,095,730	
Total value:	7,447,097,377	
Number of left outliers:	**0**	
Total quantity:		
Total value:		
Number of right outliers:	**2**	
Total quantity:	6,699	
Total value:	442,065	
Detection of outliers		
Left threshold:	0.62	
Right threshold:	32.99	
Descriptive statistics	Before removing outliers	After removing outliers
Min:	0.74	0.74
Q1:	2.75	2.75
Median:	4.98	4.93
Q3:	7.43	7.38
Max:	151.90	23.49
Arithmetic mean:	5.99	5.63
Geometric mean:	4.60	4.55
Bowley measure of skewness		
Original data:	0.04	0.06
Log-transformed data:	-0.19	-0.18
Multimodality index:	1.01	1.97

Before removing outliers — histogram of log(Unit Value) with Relative Frequency. **After removing outliers** — histogram of log(Unit Value) with Relative Frequency.

HS2 030541 - Pacific salmon /Atlantic salmon / Danube salmon [see list of conventions for s ... (Quantity unit: 8)

	Before removing outliers	After removing outliers
Number of observations:	**328**	
Total quantity:	40,093,899	
Total value:	485,711,880	
Number of left outliers:	**14**	
Total quantity:	920,284	
Total value:	3,846,846	
Number of right outliers:	**17**	
Total quantity:	61,846	
Total value:	3,020,481	
Detection of outliers		
Left threshold:	5.40	
Right threshold:	31.57	
Descriptive statistics	Before removing outliers	After removing outliers
Min:	1.67	5.53
Q1:	10.47	10.68
Median:	12.93	12.91
Q3:	16.28	15.76
Max:	315.53	27.92
Arithmetic mean:	16.24	13.46
Geometric mean:	13.12	12.85
Bowley measure of skewness		
Original data:	0.16	0.12
Log-transformed data:	0.05	0.03
Multimodality index:	1.00	1.27

Before removing outliers

After removing outliers
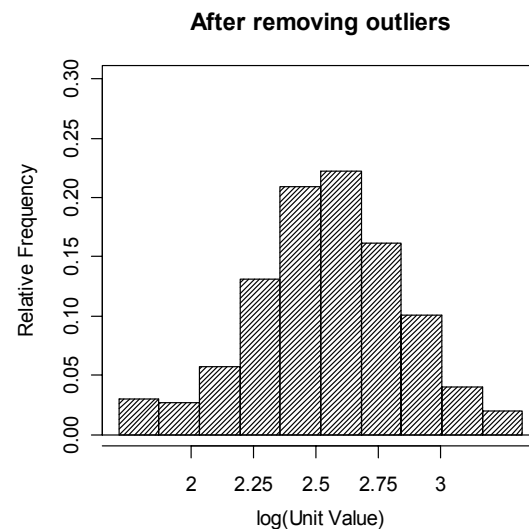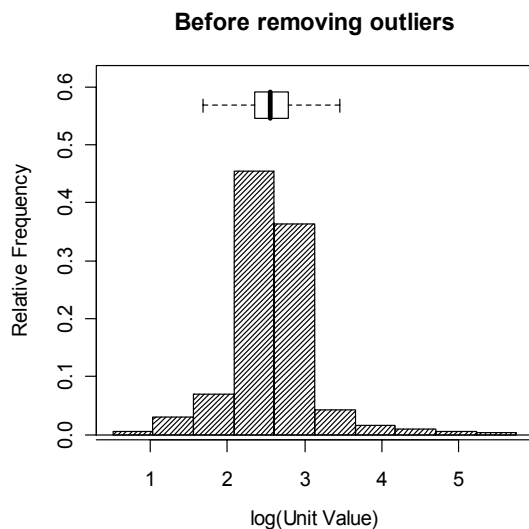
HS2 040510 - Butter
(Quantity unit: 8)

	Before removing outliers	After removing outliers
Number of observations:	**1,028**	
Total quantity:	1,093,578,404	
Total value:	3,013,485,587	
Number of left outliers:	**3**	
Total quantity:	239,628	
Total value:	143,057	
Number of right outliers:	**2**	
Total quantity:	3,175	
Total value:	53,801	
Detection of outliers		
Left threshold:	0.73	
Right threshold:	8.44	
Descriptive statistics	Before removing outliers	After removing outliers
Min:	0.55	0.80
Q1:	1.82	1.83
Median:	2.38	2.38
Q3:	3.37	3.37
Max:	143.92	7.77
Arithmetic mean:	2.80	2.66
Geometric mean:	2.48	2.48
Bowley measure of skewness		
Original data:	0.28	0.29
Log-transformed data:	0.14	0.14
Multimodality index:	1.00	2.00

Before removing outliers

After removing outliers
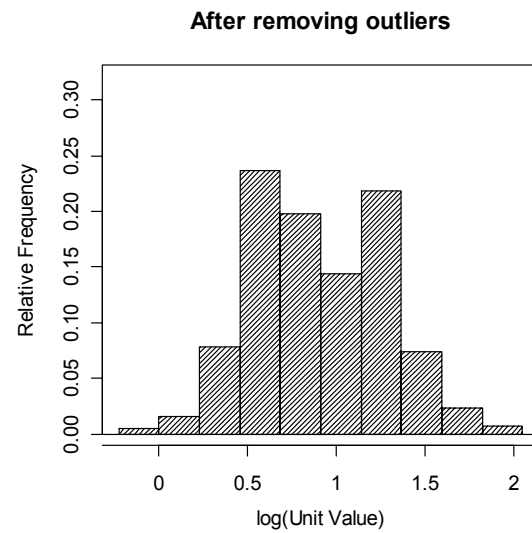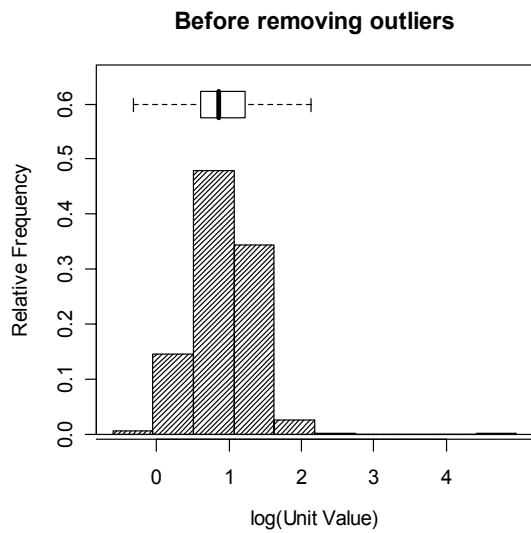
HS2 070410 - Cauliflowers & headed broccoli, fresh/chilled
(Quantity unit: 8)

Number of observations:	**243**	
Total quantity:	952,325,382	
Total value:	632,506,089	
Number of left outliers:	**16**	
Total quantity:	205,381,754	
Total value:	37,441,669	
Number of right outliers:	**4**	
Total quantity:	1,281,087	
Total value:	4,165,206	
Detection of outliers		
Left threshold:	0.24	
Right threshold:	2.87	
Descriptive statistics	Before removing outliers	After removing outliers
Min:	0.02	0.25
Q1:	0.61	0.65
Median:	0.80	0.82
Q3:	1.13	1.14
Max:	5.88	2.79
Arithmetic mean:	0.95	0.94
Geometric mean:	0.77	0.85
Bowley measure of skewness		
Original data:	0.26	0.28
Log-transformed data:	0.11	0.14
Multimodality index:	1.04	1.21

Before removing outliers

After removing outliers
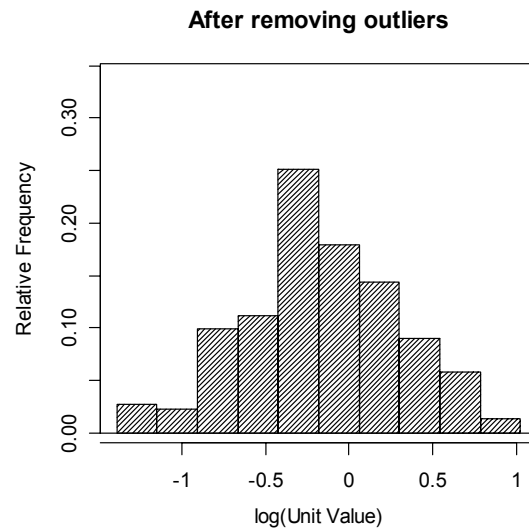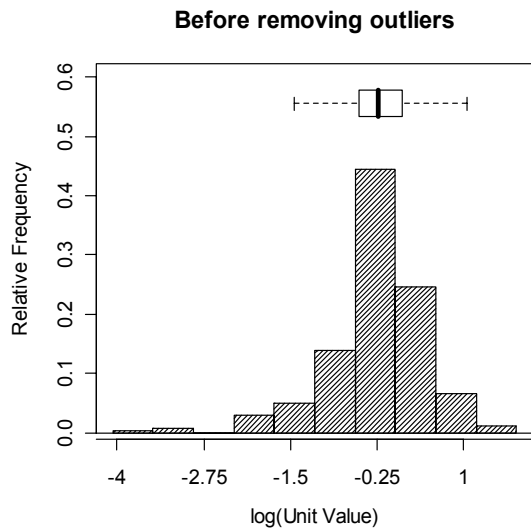
HS2 293621 - Vitamins A & their derivs.
(Quantity unit: 8)

Number of observations:	**279**	
Total quantity:	6,977,238	
Total value:	234,629,381	
Number of left outliers:	**1**	
Total quantity:	26,685	
Total value:	63,766	
Number of right outliers:	**43**	
Total quantity:	6,075	
Total value:	69,037,863	
Detection of outliers		
Left threshold:	2.40	
Right threshold:	399.28	
Descriptive statistics	Before removing outliers	After removing outliers
Min:	2.40	2.54
Q1:	16.35	14.83
Median:	26.91	22.78
Q3:	58.69	34.82
Max:	74,653.00	298.38
Arithmetic mean:	3,862.13	33.63
Geometric mean:	66.83	24.06
Bowley measure of skewness		
Original data:	0.50	0.20
Log-transformed data:	0.22	-0.01
Multimodality index:	1.42	1.14

Before removing outliers **After removing outliers**
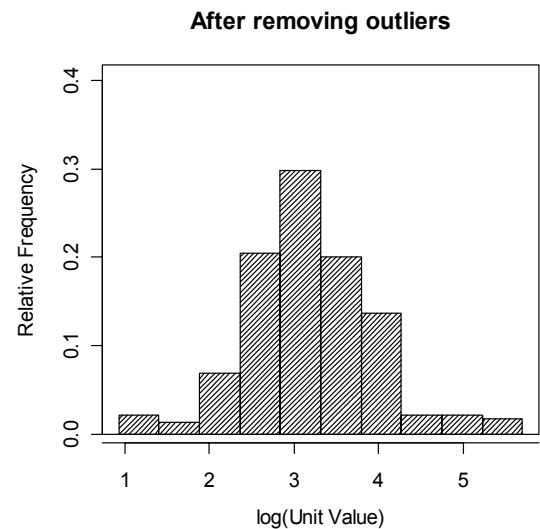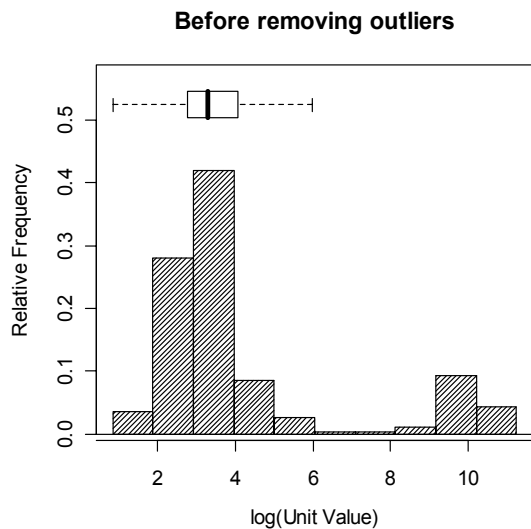
HS2 300650 - First-aid boxes & kits
(Quantity unit: 8)

Number of observations:	**284**	
Total quantity:	54,993,562	
Total value:	98,995,317	
Number of left outliers:	**1**	
Total quantity:	49,674,640	
Total value:	14,485,723	
Number of right outliers:	**42**	
Total quantity:	593	
Total value:	5,747,040	
Detection of outliers		
Left threshold:	0.95	
Right threshold:	816.77	
Descriptive statistics	Before removing outliers	After removing outliers
Min:	0.29	2.73
Q1:	11.95	11.49
Median:	19.62	16.14
Q3:	64.76	28.52
Max:	82,118.00	462.89
Arithmetic mean:	2,488.69	31.31
Geometric mean:	50.74	19.72
Bowley measure of skewness		
Original data:	0.71	0.45
Log-transformed data:	0.41	0.25
Multimodality index:	1.47	1.79

Before removing outliers

After removing outliers
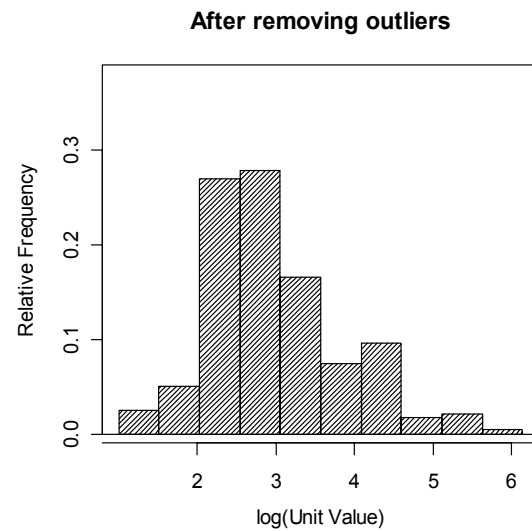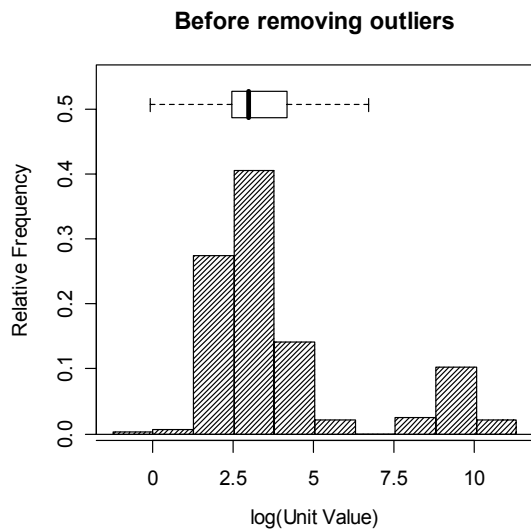
## HS2 401110 - New pneumatic tyres, of rubber, of a kind used on motor cars (incl. station ... (Quantity unit: 5)

|		
Number of observations:	**635**	
Total quantity:	112,611,567	
Total value:	3,698,809,852	
Number of left outliers:	**24**	
Total quantity:	20,413,283	
Total value:	54,237,151	
Number of right outliers:	**13**	
Total quantity:	16,255	
Total value:	4,476,692	
Detection of outliers		
Left threshold:	10.37	
Right threshold:	148.27	
Descriptive statistics	Before removing outliers	After removing outliers
Min:	1.97	11.48
Q1:	28.12	29.52
Median:	41.20	41.54
Q3:	54.68	53.80
Max:	22,907.00	145.81
Arithmetic mean:	88.63	46.03
Geometric mean:	39.14	41.00
Bowley measure of skewness		
Original data:	0.01	0.01
Log-transformed data:	-0.15	-0.13
Multimodality index:	1.11	1.00

Before removing outliers   **After removing outliers**
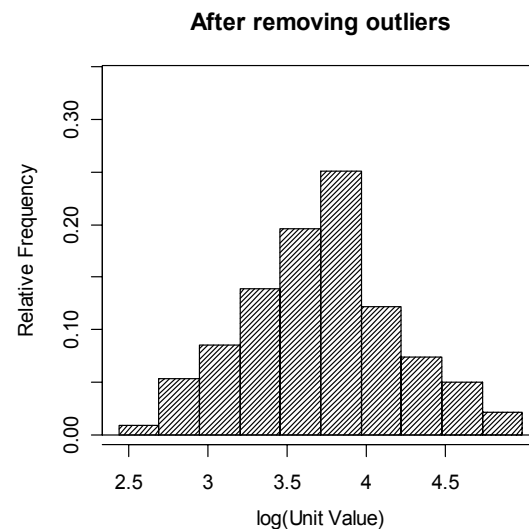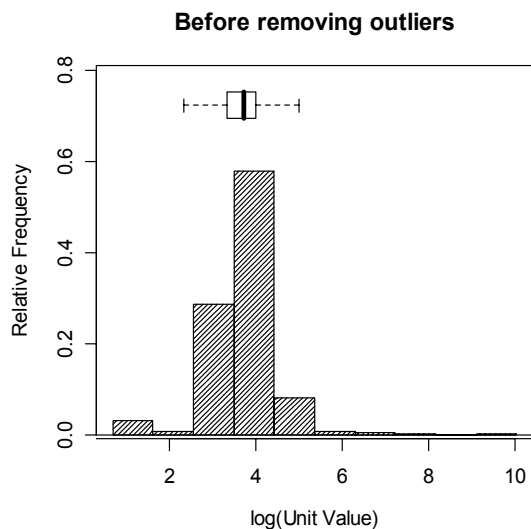
HS2 410120 - Whole bovine (incl. buffalo)/equine hides & skins, wt. per skin not >8kg... (Quantity unit: 8)

	Before removing outliers	After removing outliers
Number of observations:	399	
Total quantity:	274,503,441	
Total value:	557,433,184	
Number of left outliers:	4	
Total quantity:	6,661,475	
Total value:	803,266	
Number of right outliers:	20	
Total quantity:	164,393	
Total value:	4,174,897	
Detection of outliers		
Left threshold:	0.21	
Right threshold:	18.05	
Descriptive statistics	Before removing outliers	After removing outliers
Min:	.07	0.21
Q1:	1.12	1.10
Median:	1.88	1.80
Q3:	3.40	3.00
Max:	91.23	17.93
Arithmetic mean:	4.04	2.51
Geometric mean:	2.01	1.79
Bowley measure of skewness		
Original data:	0.33	0.26
Log-transformed data:	0.06	0.02
Multimodality index:	1.20	1.00

Before removing outliers

After removing outliers
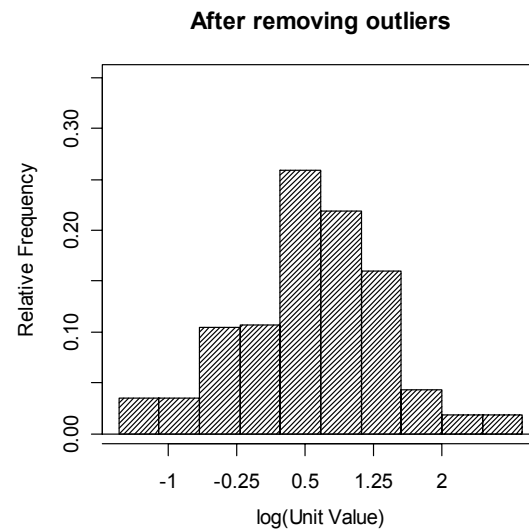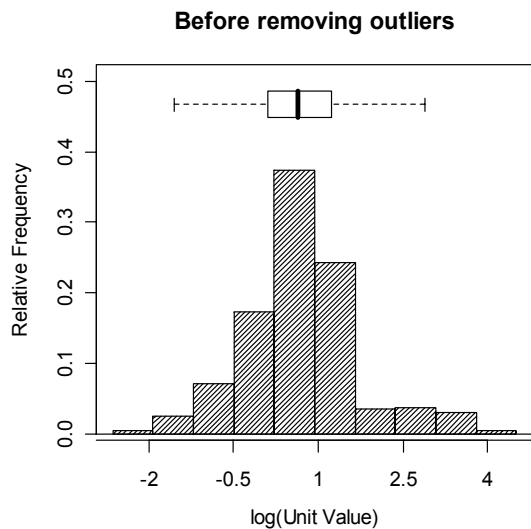
HS2 441213 - Plywood consisting solely of sheets of wood, each ply not >6mm thkns., ... (Quantity unit: 12)

Number of observations:	**17**	
Total quantity:	709,018	
Total value:	7,153,689	
Number of left outliers:	**0**	
Total quantity:		
Total value:		
Number of right outliers:	**0**	
Total quantity:		
Total value:		
Detection of outliers		
Left threshold:	0.03	
Right threshold:	343,672.68	
Descriptive statistics	Before removing outliers	After removing outliers
Min:	0.23	0.23
Q1:	12.04	12.04
Median:	364.46	364.46
Q3:	729.13	729.13
Max:	1,445.83	1,445.83
Arithmetic mean:	393.81	393.81
Geometric mean:	98.81	98.81
Bowley measure of skewness		
Original data:	0.02	0.02
Log-transformed data:	-0.66	-0.66
Multimodality index:	2.13	2.13

Before removing outliers   **After removing outliers**
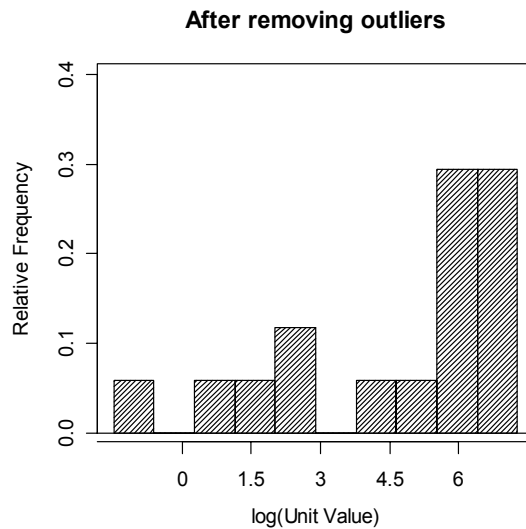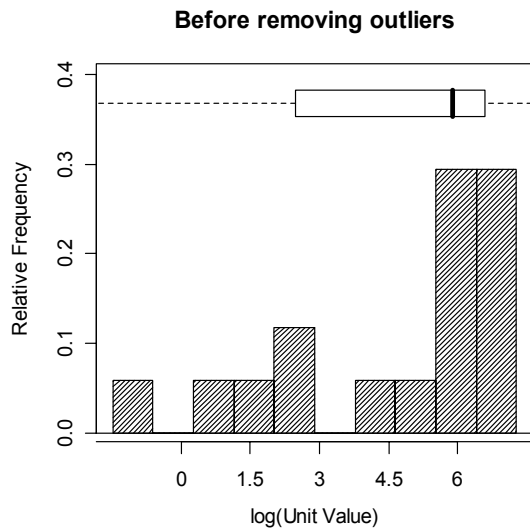
# HS2 490110 - Printed books, brochures, leaflets & sim. printed matter, in single sheets, ... (Quantity unit: 8)

Number of observations:	**1,220**	
Total quantity:	399,887,921	
Total value:	1,872,152,114	
Number of left outliers:	**20**	
Total quantity:	133,591,219	
Total value:	10,080,179	
Number of right outliers:	**48**	
Total quantity:	52,257	
Total value:	76,229,404	
Detection of outliers		
Left threshold:	0.30	
Right threshold:	322.67	
Descriptive statistics	Before removing outliers	After removing outliers
Min:	0.04	0.32
Q1:	4.14	4.18
Median:	10.46	10.01
Q3:	23.65	22.33
Max:	179,576.00	314.62
Arithmetic mean:	601.84	20.19
Geometric mean:	11.63	9.93
Bowley measure of skewness		
Original data:	0.35	0.36
Log-transformed data:	-0.06	-0.04
Multimodality index:	1.00	1.00

Before removing outliers

After removing outliers
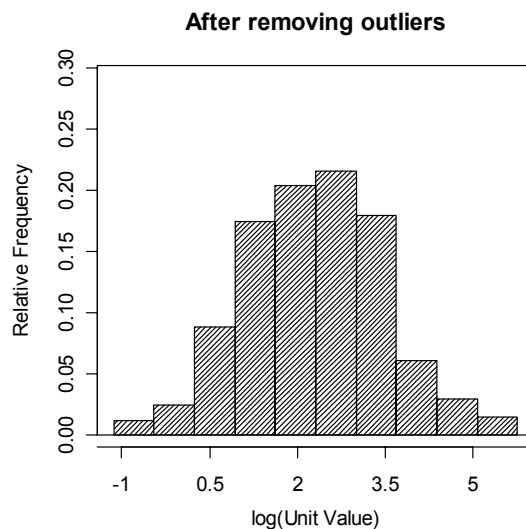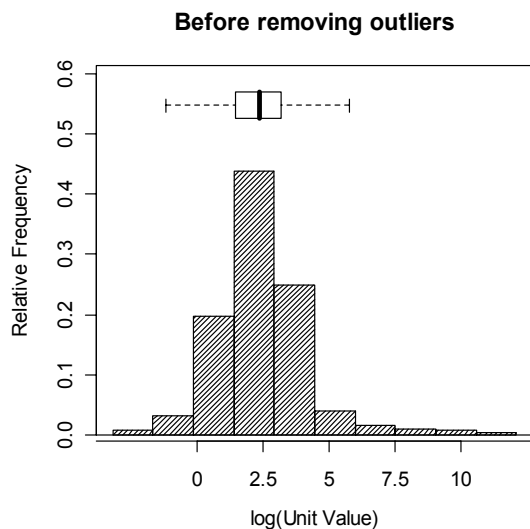
HS2 660191 - Umbrellas & sun umbrellas (excl. of 6601.10), having a telescopic shaft (Quantity unit: 5)

Number of observations:	**97**	
Total quantity:	72,950,786	
Total value:	115,820,789	
Number of left outliers:	**0**	
Total quantity:		
Total value:		
Number of right outliers:	**4**	
Total quantity:	55,747	
Total value:	1,920,686	
Detection of outliers		
Left threshold:	0.18	
Right threshold:	19.12	
Descriptive statistics	Before removing outliers	After removing outliers
Min:	0.50	0.50
Q1:	1.04	1.04
Median:	1.82	1.71
Q3:	3.33	3.16
Max:	169.79	19.03
Arithmetic mean:	5.68	2.91
Geometric mean:	2.23	1.94
Bowley measure of skewness		
Original data:	0.32	0.37
Log-transformed data:	0.03	0.11
Multimodality index:	1.27	1.29

Before removing outliers

After removing outliers
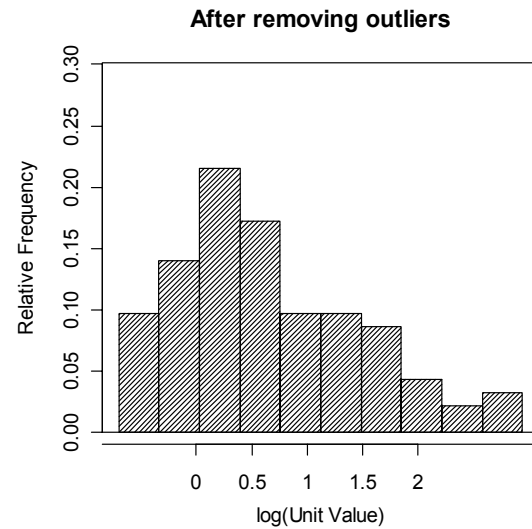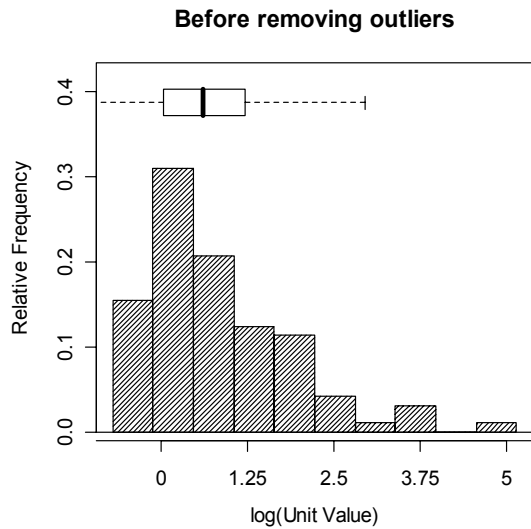
HS2 710811 - Gold (incl. gold plated with platinum), non-monetary, in powder form (Quantity unit: 8)

Number of observations:	**49**	
Total quantity:	39,743	
Total value:	195,507,576	
Number of left outliers:	**7**	
Total quantity:	22,011	
Total value:	987,910	
Number of right outliers:	**0**	
Total quantity:		
Total value:		

Detection of outliers		
Left threshold:	3,441.79	
Right threshold:	30,324.66	

Descriptive statistics	Before removing outliers	After removing outliers
Min:	12.76	5,364.07
Q1:	7,783.30	9,754.26
Median:	11,651.00	12,138.65
Q3:	13,409.62	13,562.75
Max:	17,211.73	17,211.73
Arithmetic mean:	10,019.90	11,568.40
Geometric mean:	6,750.46	11,138.06
Bowley measure of skewness		
Original data:	-0.37	-0.25
Log-transformed data:	-0.48	-0.33
Multimodality index:	1.29	1.47

Before removing outliers

After removing outliers
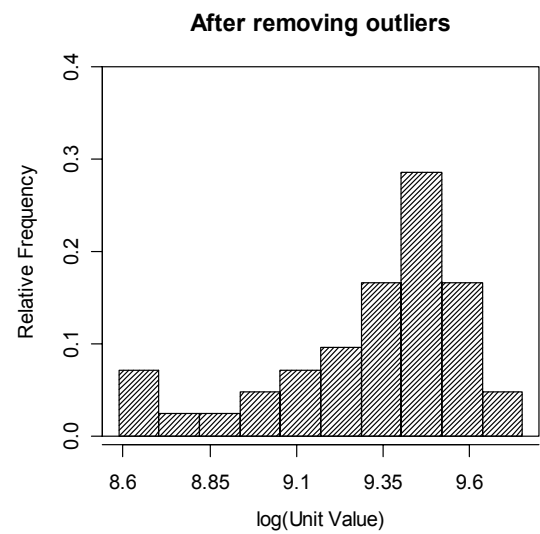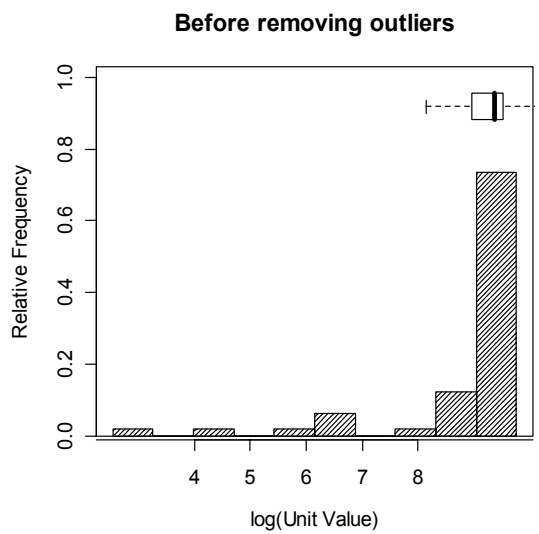
HS2 820540 - Screwdrivers
(Quantity unit: 8)

	Before removing outliers	After removing outliers
Number of observations:	**409**	
Total quantity:	38,456,377	
Total value:	226,180,751	
Number of left outliers:	**6**	
Total quantity:	20,492,479	
Total value:	1,383,272	
Number of right outliers:	**5**	
Total quantity:	27,559	
Total value:	9,622,629	
Detection of outliers		
Left threshold:	0.97	
Right threshold:	201.25	
Descriptive statistics	Before removing outliers	After removing outliers
Min:	0.05	1.41
Q1:	7.19	7.33
Median:	16.79	16.85
Q3:	27.27	27.17
Max:	2,527.30	197.01
Arithmetic mean:	30.20	21.08
Geometric mean:	14.28	14.80
Bowley measure of skewness		
Original data:	0.04	0.04
Log-transformed data:	-0.27	-0.27
Multimodality index:	1.06	1.91
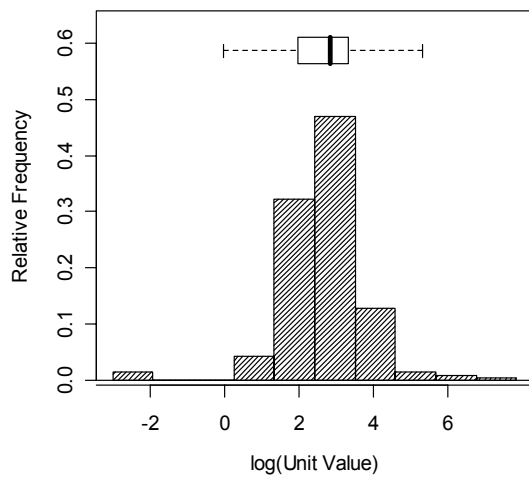
Before removing outliers   **After removing outliers**
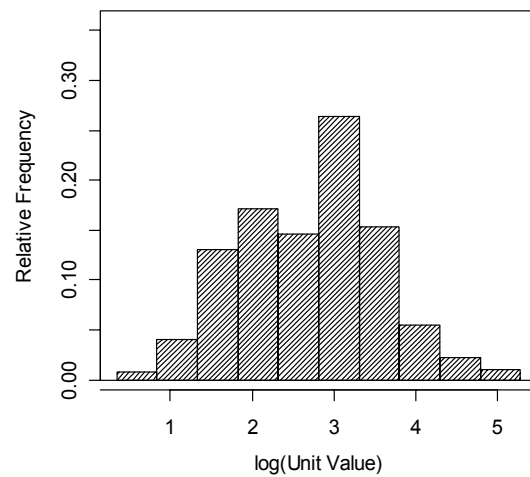
3. Standard Unit Values

- The Standard Unit Value (SUV) of a specific commodity/flow/year is defined as the **median unit value (after removing outliers).**

- Input data is taken from Tariff Line Data for those countries that have been published to UN Comtrade. For non-weight SUV, quantity is taken from supplementary units reported by countries and for weight, instead of supplementary units, reported net weight is used.

- To improve reliability, tariff line data must fulfills the following criteria:

  1. Trade value must be greater than 25000

  2. Net weight / Quantity must be greater than 0

  3. Partner countries must be individual countries not areas, such as world

  4. Net weight / Quantity must be reported as is, not estimated

- However, this is considered to be reliable for estimation purposes if and only if the sample of unit values on which it is based fulfills the following **reliability criteria**:

  5. The data must come from more than two reporting countries/regions.

  6. There must be at least 30 observations in the sample

  7. The relative standard deviation must be less than or equal to 1.75, or it must be between 1.75 and 3, provided that its multimodality index is less than 2

  8. The relative interquartile range must be less than 2

  9. The trade value corresponding to outliers must be less than 10% of the total trade value.

- The resulting Standard Unit Values for different classifications are available in the table views SuvH0, SuvH1, and SuvH2 of the StandardUnitValues data base of the UNSD.

- Standard Unit Values can be generated by executing stored procedure: pGenerateSUV