E. Estimation and imputation of quantity data

15.18.   In the absence of collected quantity information, it is good practice to estimate and impute quantity data and reflect the estimation and imputation methods in the metadata.

15.19.   *Estimation of quantity data: current practice in Germany.* Missing or erroneous quantity data are generally estimated. Quantity data are regarded as erroneous and are replaced by an estimated value if the ratio between statistical value and quantity is outside a valid range defined separately for each commodity code. In the case where an error is detected, as a general rule, it is assumed that the declared quantity (not the declared value, which is assumed to be more reliable) has led to the mistake. The acceptable range is being reviewed at least once a year and updated if necessary. The estimation is based on the average values per quantity unit. These values are calculated for each commodity code empirically with the help of plausible data relating to the preceding 12 months. The average values are updated permanently. For some commodities, a supplementary unit in addition to the net mass (e.g., the metre or the litre) is used for measuring quantity. If available, the supplementary quantity unit (instead of the unit value) is used to estimate the net mass on the basis of specific conversion factors. It must be kept in mind that the estimation of quantities may be difficult if the composition of a commodity group is heterogeneous and the unit values show a broad distribution. Hence, estimations carried out automatically should be checked manually, at least in cases of high values.

15.20.   *Unit value editing and quantity imputation: the experience of Canada*. Prior to the advent of the current editing system, Statistics Canada employed a parameter-based approach in which calculated unit values were compared with expected high and low unit values. However, this approach had several limitations: (a) given the number of classification codes, it became increasingly difficult to maintain an up-to-date set of parameters for each code; (b) although the Harmonized System provides a very detailed product classification, numerous codes include goods that are not homogeneous, resulting in extremely wide parameter sets. Consequently, a new methodology, referred to as "clipping", was developed. Essentially, this approach is based on the assumption that the majority of transactions are reported correctly and that only the outliers require correction or imputation. For each classification code, the clipping system calculates parameter sets based on the current data received. Outliers are then moved towards the mean through imputation of a corrected quantity. The principle advantages of this system are: (a) the dynamic parameters are based on more current prices; (b) the effects of seasonality are at least partially compensated for; and (c) it is far less resource-consuming.

15.21.   *Estimation methods used by UNSD for UN Comtrade*. Estimation of quantity and net weight is performed in either of two cases: where the data have not been provided or where the data provided do not conform with, and cannot be mathematically converted to, the WCO recommended quantity units. To take the best possible advantage of the information provided by a country, the quantity estimation is applied in the following sequence: 1. Estimation using empirical conversion factors; 2. Estimation using partially reported quantity and/or net weight and 3. Estimation using standard unit values. However, broad-based conversions and estimation of quantity at the national or international level are inaccurate by definition and can serve the purpose only of making quantity (especially weight) estimates for general trade or transport analyses. Estimates of quantities are sometimes also needed to preserve aggregated quantity information at the heading level of the HS.