Data compilation

The data for each surveyed variable from all enterprises producing a PRODCOM code are aggregated in order to obtain data per PRODCOM code.

More specifically:

$$PQ_p = \sum_{i=1}^{N_p} PQ_p^i$$

$$SQ_p = \sum_{i=1}^{N_p} SQ_p^i$$

$$SV_p = \sum_{i=1}^{N_p} SV_p^i$$

where:

PQ_p: is the produced quantity of PRODCOM code p,

SQ_p: is the sold quantity of PRODCOM code p,

SV_p: is the sales value of PRODCOM code p,

PQ_p^i: is the produced quantity of PRODCOM code p of enterprise i,

SQ_p^i: is the sold quantity of the PRODCOM code p of enterprise i,

SV_p^i: is the sales value of PRODCOM code p of enterprise i and

N_p: is the number of enterprises that have produced - sold PRODCOM code p.

Imputation methods

Non-response is the failure of a sample survey (or a census) to collect data for all data items in the survey questionnaire from all the population units designated for data collection. The difference between the statistics compiled on the basis of the collected data and those that would be compiled if there were no missing values is the non-response error. The two types of on-response are:

- **Unit non-response**, which occurs when no data are collected for a population unit designated for data collection.
- **Item non-response**, which occurs when data only on some, but not all the survey data items, are collected for a designated population unit.

In order to estimate the missing values the following methods are implemented.

A. Unit non-response

In the case of unit non-response, the ratio estimator for businesses producing a specific product is used, which is calculated as follows:

$$R_p = \frac{\sum_{i=1}^{N_p} SV_{p,t}^i}{\sum_{i=1}^{N_p} SV_{p,t-1}^i}$$

where:

R_p: is the sales value growth rate of product p,

$\sum_{i=1}^{N_p} SV_{p,t}^i$: is the sum of the sales value of all enterprises $(i = 1,2, \dots, N_p)$ in the current period for product p,

$\sum_{i=1}^{N_p} SV_{p,t-1}^i$: is the sum of sales value of all corresponding enterprises in the previous year $(t-1)$ for product p.

The previous period $(t-1)$ sales value of the non-responding enterprise are then multiplied by the aforementioned growth factor R_p.

Subsequently, in order to estimate the sold quantity, the sales value is divided by the unit value of product p, of the specific enterprise during the previous period $(t-1)$. Finally, the produced quantity is estimated on the basis of the the estimated sold quantity.

B. Item non-response

If the data either on sales value or sales volume for the responding enterprise k are not available, then these data are estimated via the following methods:

I. The Unit Value method at enterprise level

The estimation of the sales volume and of the sales value of product p is based on the previous period **unit value** $\left(UV_{p,t-1}^{k}\right)$, of the respondent k. The unit value is calculated as follows:

$$UV_{p,t-1}^{k} = \frac{SV_{p,t-1}^{k}}{SQ_{p,t-1}^{k}}$$

where:

$UV_{p,t-1}^{k}$: is the unit value of product p of enterprise k during period $(t-1)$,

$SV_{p,t-1}^{k}$: is the previous period sales value of the respondent k, for product p

$SQ_{p,t-1}^{k}$: is the previous period sales volume of the respondent k, for product p

Then the sales volume of product p of enterprise k is calculated as follows:

$$SQ_{p,t}^{k} = \frac{SV_{p,t}^{k}}{UV_{p,t-1}^{k}}$$

Respectively, the sales value of product p of enterprise k is calculated as follows:

$$SV_{p,t}^{k} = UV_{p,t-1}^{k} * SQ_{p,t}^{k}$$

II. The Unit Value method at product level

This method is used when the previous year unit value of the respondent k cannot be calculated. In this case, the overall unit value of respondents making product p, with both volume and value sales figures available for the current period, is calculated as follows:

$$UV_{p,t} = \frac{\sum_{i=1}^{N_p} SV_{p,t}^{i}}{\sum_{i=1}^{N_p} SQ_{p,t}^{i}}$$

where:

$UV_{p,t-1}$: is the unit value of product p during period $(t-1)$

$\sum_{i=1}^{N_p} SV_{p,t}^{i}$: is the total sales value of all responding enterprises during period t,

$\sum_{i=1}^{N_p} SQ_{p,t}^{i}$: is the total sales volume of all responding enterprises during period t.

N_p: all the respondents making product p, with both sales volume and sales value figures available for the current period t.

It should be noted, that any outlier is removed before calculating the unit value.

The aforementioned unit value is then used to estimate either the sales value or sales volume of the respondent k, depending on the data availability.

It is clear that these practices cannot be implemented for enterprises which are included in the survey for the first time.