Labour Cost Survey

TYPE
   Sampling survey, which concerns the total labour cost of the enterprises

Sample design
The labour costs survey covers the sections C-K and M-O of the NACE Rev.1 and the enterprises with average annual employment equal to or greater than 10 employees. The single stratified random sampling method was applied, employing the enterprise as a surveyed unit and obtaining statistical information from *each separate local unit* of the enterprises included in the sample.
 The sampling frame used for the sample design was based on the Business Register (BR) of the NSSG. This BR is based on the VAT Register of the Ministry of Economy and Finance and it is updated through the statistical surveys of the NSSG and the register of the Social Insurance Foundation.
The statistical data for the public services of the sections M (Ministry of education and public schools) and N (Ministry of Health, public hospitals and public health centres) were collected from the Ministry of National Education and Religion Affairs, and the Ministry of Health and Social Solidarity, respectively.

Stratification
The enterprises with 10 or more employees included in the survey were stratified as following:
   a. By geographical region – NUTS I,
   b. By Division (two – digit NACE Rev.1 code) within each geographical region (Geography x Economic activity = Major stratum), and
   c. By size class of the enterprise. In each of the major strata, the enterprises were stratified into H=7 size classes, according to their size, determined by their average annual number of employees in the business register, as follows.

Class 1	10-19	Employees
Class 2	20-49	"
Class 3	50-99	"
Class 4	100-249	"
Class 5	250-499	"
Class 6	500-999	"
Class 7	1000 or more	"

The enterprises that belong to the 7th size class were surveyed exhaustively.

Sample size
The sample size is 4.313 enterprises (sampling fraction=20,1%). The sample size of the enterprises was defined, so that the relevant standard error (co-efficient of variation CV) of the variables "annual labour costs" and "hourly labour costs" at 2-digit code level of economic activity at the whole country does not exceed 5%. The sampling units (enterprises) were distributed to size strata applying the method of optimal (Neyman) allocation.

Selection of the sampling units (enterprises)

In each of the final strata (let h), a sample of n_h enterprises was selected. The enterprises to be surveyed were selected from the total of the N_h enterprises with equal probabilities and by applying systematic sampling. The sampling units (enterprises) were selected from the sample frame based on data from Business Register of the NSSG.

Survey characteristics estimation
a. Symbols
Defining with index *i* the selection order of an enterprise from the sampling frame in the stratum *h* and symbolizing with the *y* one of the survey characteristics, we can define the following:

y_{hi} : the value of the survey characteristic *y* of the enterprise of order *i* in the stratum *h*

Y_h : the sum of the values of the characteristic *y* for all enterprises falling into the survey and belonging to the stratum *h*

Y : the sum of the values of the characteristic *y* of all enterprises under survey belonging to one economic activity with two digit code . That is:

$$Y = \sum_h Y_h$$

b.Estimation process
The estimations of the magnitudes Y_h and *Y* come from the following relations:

$$\hat{Y}_h = \frac{N_h}{n_h} \cdot \sum_{i=1}^{n_h} y_{hi}$$

$$\hat{Y} = \sum_h \hat{Y}_h$$

Generally, in order the estimations of the survey characteristics to be produced at any level, we add up the estimations of the (final) strata, which form the level under survey. The estimates of totals are produced using the Horvitz-Thompson estimator of the first relation above, which is *unbiased*.

There are quantities being produced through the ratio of two variables (as the annual labour costs per employee or the hourly labour costs). The estimations are produced using the ratio estimator, which is usually *slightly biased*.

We assume that the population parameter to be estimated is the ratio:

$$R = \frac{\sum_{i=1}^{N} y_i}{\sum_{i=1}^{N} x_i} = \frac{Y}{X} = \frac{\bar{Y}}{\bar{X}}$$

where y_i and x_i are the values for each unit of order *i* in the population of size *N* (e.g. the variable *y* is the total labour cost and the variable *x* is the number of hours actually worked). If the stratified random sampling scheme is applied then the combined estimation of R is:

$$\hat{R} = \frac{\sum_h \frac{N_h}{n_h} \sum_{i=1}^{n_h} y_{hi}}{\sum_h \frac{N_h}{n_h} \sum_{i=1}^{n_h} x_{hi}} = \frac{\hat{Y}}{\hat{X}}$$

c. Variance estimation

The estimations of the variances of $\hat{Y}_h$ and $\hat{Y}$ come from the following relations:

$$Var(\hat{Y}_h) = \frac{N_h(N_h - n_h)}{n_h(n_h - 1)} \cdot \left[\sum_{i=1}^{n_h} y_{hi}^2 - \frac{\left(\sum_{i=1}^{n_h} y_{hi}\right)^2}{n_h} \right]$$

$$Var(\hat{Y}) = \sum_h V(\hat{Y}_h)$$

The coefficient of variation (%) of the $\hat{Y}$ is given by the following relation:

$$CV(\hat{Y}) = \frac{\sqrt{Var(\hat{Y})}}{\hat{Y}} \cdot 100$$

The estimation of the variance of $\hat{R}$ is calculated from the following relation:

$$Var(\hat{R}) = \frac{1}{\hat{X}^2} \cdot \sum_h \frac{N_h \cdot (N_h - n_h)}{n_h} \cdot \left[S_{yh}^2 + R^2 \cdot S_{xh}^2 - 2 \cdot R \cdot Cov(y_h, x_h) \right]$$

where: $S_{yh}^2 = \frac{1}{n_h - 1} \cdot \left[\sum_{i=1}^{n_h} y_{hi}^2 - \frac{\left(\sum_{i=1}^{n_h} y_{hi}\right)^2}{n_h} \right]$, $S_{xh}^2 = \frac{1}{n_h - 1} \cdot \left[\sum_{i=1}^{n_h} x_{hi}^2 - \frac{\left(\sum_{i=1}^{n_h} x_{hi}\right)^2}{n_h} \right]$,

$$Cov(y_h, x_h) = \frac{1}{n_h - 1} \cdot \left[\sum_{i=1}^{n_h} y_{hi} \cdot x_{hi} - \frac{\left(\sum_{i=1}^{n_h} y_{hi}\right) \cdot \left(\sum_{i=1}^{n_h} x_{hi}\right)}{n_h} \right]$$

The coefficient of variation of $\hat{R}$ is calculated from the following relation:

$$CV(\hat{R}) = \frac{\sqrt{Var(\hat{R})}}{\hat{R}} = \frac{1}{\hat{Y}} \cdot \sqrt{\sum_h \frac{N_h \cdot (N_h - n_h)}{n_h} \cdot \left[S_{yh}^2 + R^2 \cdot S_{xh}^2 - 2 \cdot R \cdot Cov(y_h, x_h) \right]}$$

The coefficient of variation of the variables "annual labour costs" and "hourly labour costs" are shown in the *Annex I*, according to the structure of tables A (national data), B (regional data) and C (national data by size class of enterprise).

In the section with code E, the coefficient of variations of the variables "annual labour costs" and "hourly labour costs" are equal to 0,1%, because 90% of the total statistical information was collected from two large enterprises (average annual employment higher than 1000 persons) belonging to the census (take –all) stratum.

In the sections M and N, the coefficient of variations of the variables "annual labour costs" and "hourly labour costs" are less than 1%, because the statistical information for the public services was collected from administrative sources (Ministry of National Education and Religion Affairs, Ministry of Health and Social Solidarity).

In the sections with codes D, I and J, the coefficient of variations of the variables "annual labour costs" and "hourly labour costs" are less than 1,6%.

In the sections with codes C, G, K and O, the coefficient of variations of the variables "annual labour costs" and "hourly labour costs" are ranged between 3% and 5%.

In the sections with codes F and H, the coefficient of variation *only* of the variable "annual labour costs" is higher than 5%, because in these sections strong seasonality is appeared. As a result, two different types of enterprises belong to the same size classes, as follows:

- Enterprises operating all the year
- Enterprises operating only a time of period less than one year (approximately, half a year)

The enterprises of the first type have annual labour costs higher than the enterprises belong the second type, and as a result, in the same size stratum the annual labour costs of the enterprises are not homogeneous due to the different types of enterprises. Thus, although in the same size stratum, internally homogenous enterprises exist, according to their number of employees, however internally heterogeneous enterprises are appeared, according to their values of the "total annual labour costs". This increases the variance of the total annual labour costs of the enterprises, reducing the gain in the precision from the stratification that was introduced initially in the sample selection.

Concerning, the problem of the internal heterogeneity in the size strata was not appeared in the variable "hourly labour costs", because the variables "annual labour costs" and "hours actually worked" are strong correlated. As a result, there was not any high variability in the hourly labour costs due to the different types of enterprises.

ANNEX I

The coefficients of variations (%) of the quantities "annual labour costs" and "hourly labour costs" are appeared broken down according to the structure of the tables A, B and C.

Table A. Coefficient of variations (%) by economic activity (Sections)

NACE Rev. 1	Annual labour costs	Hourly labour costs
Total	0,8	0,7
C	4,5	3,8
D	1,4	1,0
E	0,1	0,1
F	6,1	4,1
G	3,3	3,2
H	7,5	5,0
I	1,5	1,2
J	1,0	1,0
K	4,8	4,5
M	0,2	0,2
N	0,6	0,4
O	4,9	4,2

Table B. Coefficient of variations (%) by NUTS level 1

NUTS1	Annual labour costs	Hourly labour costs
Total	0,8	0,7
North Greece	1,7	1,4
Central Greece	1,9	1,4
Attica	0,9	0,8
Islands of Aegean and Crete	4,4	2,8

Table C. Coefficient of variations (%) by size class of enterprises

Size classes	Annual labour costs	Hourly labour costs
Total	0,8	0,7
10-49	3,3	2,4
50-249	1,2	1,0
250-499	2,1	1,7
500-999	2,7	1,9
1.000+	0,0	0,0