INSTITUTO NACIONAL DE ESTADISTICA

Labour Price Index (IPT)

Methodology

Madrid, 2020.

Contents

1 Objectives

A Labour Price Index (IPT) is an indicator that aims to measure the change in the price of labour over time as a direct consequence of market pressures, that is, without this measurement being affected by changes in the quality and quantity of work done (for example, by changes in the composition of the workforce, the number of hours worked, the type of contract, the characteristics of the employees, etc.) or, in other words, discounting the composition effect[1].

By analogy with the Consumer Price Index, which measures changes in the price of a standardised *basic basket* of consumer goods and services, the IPT must measure the changes in the price of a standardised *basic basket* of jobs.

To measure the change in the pure price of a group of items between two periods, the following conditions must be met: first, the set of basic elements must be the same, and secondly, the characteristics of all the items must remain unchanged. These two conditions applied to the present case require a highly detailed study of jobs.

The best way to develop a labour price index is by organising a specific survey, usually temporary, similar to that in the US (Employment Cost Index, ECI), in which data on the jobs included in the*basic basket* are generally taken quarterly.

The organisation of such a survey is expensive, so at the present time this is not feasible in Spain. Therefore, the decision has been made to conduct the annual monitoring of a sample of work positions which are representative of those in the overall economy, drawing on information provided by existing surveys, that is, without an additional cost for collecting information.

[1]The information currently existing in Spain, from the Quarterly Labour Cost Survey, the Annual Labour Cost Survey and Structure of Earnings Surveys of the INE consists of series of average results of labour and wage cost per worker. These series include both the variation in cost and the changes in the composition of employment.

2 Information sources used

The Structure of Earnings Surveys is the source of information chosen for the preparation of the IPT, for the following reasons:

- They are an existing source of information.
- They offer a wealth of information: They have data on individual employees, particularly on the occupation variable, which is the central variable in thistype of analysis.
- Their sample size allows the study of the characteristics of employees in great detail.
- Although they do not offer information on all the components of labour costs, they do offer the most important: the salary. They also have the normal working hours necessary for calculating the Index.

The Structure of Earnings Survey is a four-yearly investigation (starting from 2002) into the structure and distribution of wages, carried out in all Member States of the European Union. Common criteria in methodology and content are used in order to obtain comparable results regarding the level, structure and distribution of wages among EU Member States in accordance with Regulations 530/1999 and 1738/2005.

The main new feature it provides compared to other surveys on this subject is that wages are collected in the questionnaire individually, and together with these, a large amount of variables related to the employee. Thanks to this, it is possible to establish relationships between wages and some of those variables which may contribute to determining their amount, such as the level of studies attained, seniority, type of contract or occupation, among others.

From 2004 and for the years in which the quadrennial survey is not performed, the Structure of Earnings Annual Survey obtained (for Spain, this is not a European survey) estimates of the annual gross earnings per employee classified by type of working day, sex, economic activity and occupation.

In this survey the information is obtained through the combined processing of the Social Security General Affiliation File (SS) and declarations of Model 190: Annual Summary of Withholdings and Advance Payments on Personal Income Taxes by the State Tax Administration Agency (AEAT) and the Treasury of Comunidad Foral de Navarra, Bizkaia and Gipuzkoa, along with the employment and work time variables provided by the survey attached to the INE Quarterly Labour Cost Survey.

The annual and quadrennial surveys are designed in the same way and their fields of study, reference periods, etc. are the same. The random unit selection procedure

corresponds to a stratified two-stage sampling, in which the first stage units are the Social Security contribution accounts (the General Treasury of the Social Security provides the framework of contribution accounts), while those of the second stage are the employees.

The first stage units are stratified by economic activity, the Autonomous Community and size ranges (measured in number of employees).

After selecting the contribution accounts, the General Social Security System (TGSS) obtains the list of employees who were registered during the whole of the month of October in the reference year. Employees are selected from this list, as are the second-stage units.

In the case of the annual survey, the first-stage sample matches the sample from the Quarterly Labour Cost Survey (QLCS). Together with the issuing of the QLCS questionnaires for a specific quarter, we include an attached questionnaire with the sample of employees selected in each contribution account and identified by their Social Security number, essentially requesting information on the variable occupation.

The information from the QLCS and the attached questionnaire is cross-referenced with that from the Tax Administration (taken from the annual summary file of Withholdings and Payments on Account regarding Personal Income Tax) to incorporate the data on the annual income of each employee.

The sample sizes are about 28,000 contribution accounts and more than 200,000 employees.

The sample units of the annual survey, to match the QLCS sample, are divided into five rotation groups, in such a way that, in the first quarter of every year, the oldest group is replaced, which entails an annual renewal of 20% of the sample. However, in the years that the quadrennial survey is carried out, the sample is selected in such a way that it is not coincident with the QLCS sample except in the comprehensive section (units of more than 500 employees).

3 Scope of the statistical operation

In using the Estructure of Earnings Surveys for the development of the IPT, the population, geographical and temporal scope as well as the sectoral coverage of the Index are those derived from these surveys. They are shown below:

The population scope is made up of all those workers who work for others (employees), who provide their services in enterprises, and who have been affiliated to Social Security for more than two months of the year, including October. The choice of this month has the advantage of being considered *normal* in all EU countries, in the sense that this is not affected much by seasonal variations or by payments which fall due in more than one month?s time, such as Christmas bonuses.

It excludes all chairpersons, members of administrative boards, and in general, all personnel whose remuneration is not mainly in the form of wages, but rather commissions or benefits.

The geographical scope encompasses the entire country, with a breakdown of results by Autonomous Community. The information corresponding to Ceuta and Melilla is provided jointly with that corresponding to Andalucía.

As for the **sectorial coverage**, contribution centres are studied, regardless of their size, whose economic activity is included in the three large sectors: Industry, Construction and Services.

The survey excludes agricultural, livestock and fishing activities, domestic staff and foreign agencies.

For the **temporal scope**, the reference period is the calendar year.

4 Classification variables

The classification variables used to determine the jobs considered in the preparation of the IPT are as follows:

- **Activity:** sections of activity according to the CNAE-09 from the sampling field of the Structure of Earnings Surveys: B, C, D, E, F, G, H, I, J, K, L, M, N, O, P, Q, R and S.

  B Extractive industries

  C Manufacturing industry

  D Supply of electrical energy, gas, steam and air conditioning

  E Water supply, waste management and decontamination activities

  F Construction

  G Wholesale and retail trade; repair of motor vehicles and motorcycles

  H Transport and storage

  I Accommodation

  J Information and communications

  K Financial and insurance activities

  L Real estate activities

  M Professional, scientific and technical activities

  N Administrative and support services activities

  O Public Administration and defence; Compulsory social security activities

  P Education

  Q Health and social services activities

  R Arts, recreation and entertainment activities

  S Other services

- **Size of work centre (stratum):** takes 3 different values, depending on thenumber of employees at the centre:

  1 From 1 to 49 employees

  2 From 50 to 199 employees

  3 200 or more employees

- **Region:** Autonomous Community in which the work centre is located:

  01 Andalucía (includes Ceuta y Melilla)

  02 Aragón

  03 Asturias (Principado de)

  04 Balears (Illes)

  05 Canarias

  06 Cantabria

  07 Castilla y León

  08 Castilla - La Mancha

  09 Cataluña

  10 Comunitat Valenciana

  11 Extremadura

  12 Galicia

  13 Madrid (Comunidad de)

  14 Murcia (Región de)

  15 Navarra (Comunidad Foral de)

  16 País Vasco

  17 Rioja (La)

- **Sex:** Females and males.

- **Occupation:** the large groups from the National Classification of Occupations (CNO-11): from 1 to 9. Groups 0 and 5, and 6 and 7 appear together due to problems of sample representation in the Estructure of Earnings Surveys.

  1 Directors and managers

  2 Scientific and intellectual professionals and technicians

  3 Technicians, support professionals

  4 Accountancy, administrative and other office employees

  5 and 0 Workers in catering, personal, and protection services and salespersons Military occupations

  6 and 7 Skilled agricultural, livestock, forestry and fishing sector workers Craftspersons and skilled workers in manufacturing and construction (except installation and machinery operators)

  8 Machinery and installation operators and assemblers

  9 Elementary occupations

- **Type of contract:** permanent and temporary.
- **Contractual working time:** full-time and part-time.
- **Age of employee:** takes 5 different values, determined by the followingage ranges:

  1 Under 25 years old

  2 25-34 years old

  3 35-44 years old

  4 45-54 years old

  5 55 years old or over

- **Nationality of the employee:** Spanish and other nationalities.
- **Length of service of the employee in the workplace:** 6 different values, determined by the following intervals:

  1 less than 1 year in service

  2 1-3 years in service

  3 4-10 years in service

  4 11-20 years in service

  5 21 years or more in service

5 Calculation methodology

5.1 Index calculation formula

A chained Laspeyres index has been used for the preparation of the IPT.

It was chosen because there can be changes in the structure of employment from one year to the next and this kind of indicator allows us to enter them annually through the weightings without having to wait for a base change as occurs in the fixed base index.

The formulation of the IPT is as follows:

The letter c denotes the cross-sections of the variables that define the characteristics of the jobs (detailed in section 4). That is, what is constituted by said cross-sections is considered a fixed structure, with the aim of determining how the hourly earnings have varied while discounting the composition effect.

Being,

G_c^k hourly earnings of employees in cross-section c in the year k,

h_c^k hours worked by employees in cross-section c in the year k.

$\omega_c^k = G_c^k \cdot h_c^k$ annual earnings of employees in cross-section c in the year k.

Definition of the basic Laspeyres formula to be used to calculate the IPT during the year j, for the base year k is as follows.

$$IPT_{j(k)} = \frac{\sum\limits_c G_c^j h_c^k}{\sum\limits_c G_c^k h_c^k} = \frac{\sum\limits_c \frac{G_c^j}{G_c^k} G_c^k h_c^k}{\sum\limits_c G_c^k h_c^k} = \frac{\sum\limits_c \frac{G_c^j}{G_c^k} \omega_c^k}{\sum\limits_c \omega_c^k},$$

where $\frac{\omega_c^k}{\sum\limits_c \omega_c^k}$ are the weightings for calculating the index and $\frac{G_c^j}{G_C^k}$ re the simple or elementary indices at the cross-section c.

Definition of the link coefficient for the year j with the year $j+1$ is,

$$L_{j,j+1} = \frac{\sum\limits_c G_c^{j+1} h_c^j}{\sum\limits_c G_c^j h_c^j}.$$

Therefore, the formula for the chained Laspeyres index for the year j and with reference to the year k is,

$$IPT_{j(k)} = 100 \cdot L_{k,k+1} \cdot L_{k+1,k+2} \cdots L_{j-2,j-1} \cdot IPT_{j(j-1)}.$$

In addition, it is noted that the link coefficient $L_{j,j+1}$ is equivalent to the index for the year $j+1$ with base year the year j, that is: $L_{j,j+1} = IPT_{j+1(j)}$.

Therefore, the Laspeyres index would be as follows,

$$IPT_{j(k)} = 100 \cdot IPT_{k+1(k)} \cdot IPT_{k+2(k+1)} \cdots IPT_{j-1(j-2)} \cdot IPT_{j(j-1)},$$

The formula for calculating the chained index for the year j of the combination Q ofcross-sections c, being the base year k would be as follows,

$$IPT_{j(k),Q} = 100 \cdot IPT_{k+1(k),Q} \cdot IPT_{k+2(k+1),Q} \cdots IPT_{j-1(j-2),Q} \cdot IPT_{j(j-1),Q}.$$

5.2 Base period

The year 2016 is the base period, which means that all the indices that are calculated will refer to this year, and for this year, the value of the index is 100.

5.3 Reference period of weightings

The structure of the weightings establishes the importance of each job compared to the others based on the salary cost paid by the employer. For each year, the measuring of the weightings is done from hourly earnings and hours worked by employees associated with each type of job from the previous year.

5.4 Calculation of hourly earnings

To obtain a more robust estimate of the hourly earnings in different jobs defined by the classification variables, a hedonic regression model is used. It is not possible to incorporate the observed data directly, since there are jobs that appear/disappear from one year to the next and the sample representation of some jobs is scarce.

The regression model used in calculating the hourly earnings is as follows:

For each year j, it is assumed that the hourly earnings G for an employee who belongs to a cross-section c dof the explanatory variables in the model is:

$$l_c^j = \ln G_c^j = x_c' \beta^j + \varepsilon_c^j, \tag{1}$$

where,

x_c' is a vector of dimension $(1\text{x}p)$, whose elements are equal to 0 or 1, depending onthe characteristics that define cross-section c, in terms of the main effects and interactions,

β^j is a vector of dimension (px1), which contains the unknown parameters from the model,

ε_c^j is the random component of the model for the cell c in the year j.

The vector β^j defines the proportional effect on hourly earnings of p dichotomous variables included in x_c'. The p unknown parameters include the constant and the coefficients of the dichotomous variables associated with the main effects and the interactions of the regression model.

For each r possible categories that have a major effect, the model includes $(r-1)$ parameters and if the interaction has (rxs) possible combinations of values, the model would include $(r-1)$x$(s-1)$ parameters.

The random components verify:

$$E[\varepsilon_c^j] = 0; \quad Var[\varepsilon_c^j] = \sigma_j^2; \quad Cov[\varepsilon_c^j, \varepsilon_d^j] = 0 \ \ \forall\, c \neq d.$$

Having defined the model, the parameter vector β^j must be estimated with the information available each year. To this end, the model (1) is prepared in matrix notation, in the following manner:

$$L^j = X^j \boldsymbol{\beta}^j + \boldsymbol{\varepsilon}^j$$

where,

L^j is a vector of dimension (n_jx1) containing the n_j elements of the year j previously denoted by $l_c^j \ \forall\ j = 1, \ldots, n_j$. That is, it contains as many rows as there are employees in the sample from year j.

X^j is a matrix of dimension (n_jxp), whose elements are equal to 0 or 1. Each row represents an employee and each column contains the p characteristics that define this employee.

$\boldsymbol{\beta}^j$ is a vector of dimension (px1) containing the unknown parameters of the year j. It includes the constant and the coefficients associated with the main effects and interactions from the model.

$\boldsymbol{\varepsilon}^j$ is a vector of dimension (n_jx1) containing the n_j random components of the model in the year j. This vector verifies:

$$E[\boldsymbol{\varepsilon}^j] = \mathbf{0}; \quad Var[\boldsymbol{\varepsilon}^j] = \sigma_j^2 I_{n_j \text{x} n_j}.$$

As the data used in the regression model are derived from samples drawn from a population with a given sample design, we calculate the estimate of $\boldsymbol{\beta}^j$ using the weighted least squares method for sampling weights (MCOW):

$$\widehat{\boldsymbol{\beta}}^j = (X'^j W^j X^j)^{-1} X'^j W^j L^j,$$

where W^j is a diagonal matrix of dimension n_jxn_j containing sample weights. And its variance is:

$$Var[\widehat{\boldsymbol{\beta}}^j] = \sigma_j^2 (X'^j W^j X^j)^{-1} X'^j W^{j^2} X^j (X'^j W^j X^j)^{-1} = V^j,$$

where the matrix V^j has the dimension (pxp).

The estimated hourly earning for each cross-section is obtained from the formula (1), whereby the earnings estimated in the cross-section c in the year j come to:

$$\widehat{G}_c^j = \exp\{x_c' \widehat{\beta}^j\}. \tag{2}$$

The problem with this estimator is that it is biased, because although estimator β^j is indeed unbiased, when calculating its exponential this characteristic is lost. In order to correct this bias, the estimator proposed by El-Shaarawi and Viveros (1997) is used:

$$\widehat{G}_c^j = \exp\{x_c' \widehat{\beta}^j - \frac{1}{2} x_c' \widehat{V}^j x_c + \frac{1}{2} \widehat{\varphi}_j^2 \widehat{\sigma}_j^2\}. \tag{3}$$

The estimator (3) substantially corrects the bias of the estimator (2), assuming the normality of the errors ε_c^j .

For the estimation of the variance shown in the above expressions, the residuals e_c^j are defined as the difference between the logarithms of the observed andestimated hourly earnings:

$$e_c^j = l_c^j - x_c' \widehat{\beta}^j.$$

The varianze $\widehat{\sigma}_j^2$ is estimated with the average of the residual quadrants:

$$\widehat{\sigma}_j^2 = \frac{1}{n_j - p} \sum_c^{n_j} (e_c^j)^2.$$

Being the estimation of: φ_j^2:

$$\widehat{\varphi}_j^2 = 1 - \frac{\widehat{\sigma}_j^2}{2(n_j - p)} - \frac{\widehat{\sigma}_j^4}{3(n_j - p)^2}. \tag{4}$$

However, in (4) the variances of the numerator are less than 1 and in dividing them by the sample size they give $\widehat{\phi}_j^2 \approx 1$, so the expression (3) is finally:

$$\widehat{G}_c^j = \exp\{x_c' \widehat{\beta}^j - \frac{1}{2} x_c' \widehat{V}^j x_c + \frac{1}{2} \widehat{\sigma}_j^2\}.$$

6 Dissemination

Data for the national and autonomous region totals have been published (the Autonomous Cities of Ceuta and Melilla are provided together with Andalucia)

As for the classification variables of the jobs, disaggregated information is published by:

- Activity section
- Stratum (size of the workplace)
- Gender
- Occupation
- Type of contract
- Contractual working time
- Age
- Nationality
- Length of service

The weights used in each year disaggregated by Autonomous Community, Section of activity and size of the workplace will also be provided.