



Índice de precios del trabajo (IPT)

Proyecto Técnico

**Subdirección General de Estadísticas del Mercado Laboral
Junio de 2015**

Índice

	Página
1. Objetivos del proyecto	4
2. Características del proyecto	5
3. Metodología de cálculo	8
4. Plan de difusión y periodicidad.....	12
5. Calendario de implantación	12
6. Presupuesto estimado.....	12

1. Objetivos del proyecto

Un índice de precios del trabajo (IPT) es un indicador cuyo objetivo es medir el cambio en el precio de la mano de obra en el tiempo como consecuencia exclusivamente de las presiones del mercado laboral, es decir, sin que dicha medida esté afectada por cambios en la calidad y cantidad de trabajo realizado (por ejemplo, por cambios en la composición de la fuerza de trabajo, en el número de horas trabajadas, en las características de los asalariados, etc.) o, lo que es lo mismo, descontando el efecto composición¹.

Por analogía con el índice de precios al consumo que mide los cambios en el precio de una "cesta básica" estandarizada de bienes de consumo y servicios, el IPT debe medir los cambios en el precio de una "cesta básica" estandarizada de puestos de trabajo. Además, con el fin de medir el cambio en el precio puro de un grupo de ítems entre dos períodos, deben cumplirse las dos condiciones siguientes: en primer lugar, el conjunto de elementos básicos debe ser el mismo, y en segundo lugar, las características de todos los ítems deben mantenerse inalteradas. Estas dos condiciones exigen una caracterización de los puestos de trabajo muy detallada.

La forma más adecuada de elaborar un índice de precios del trabajo es mediante la organización de una encuesta específica, generalmente coyuntural, similar a la existente en Estados Unidos (Employment Cost Index, ECI), en la que trimestralmente se toman datos de los puestos de trabajo incluidos en la 'cesta básica'. La organización de tal encuesta tiene un coste elevado, por lo que en el momento actual es inabordable en España. Por ello, se ha optado por efectuar el seguimiento anual de una muestra de puestos de trabajo representativos de los existentes en el conjunto de la economía, aprovechando la información proporcionada por encuestas ya existentes, esto es, sin coste adicional de recogida de información.

¹ La información existente actualmente en España, procedente de la Encuesta Trimestral de Coste Laboral, la Anual de Coste Laboral y las de Estructura Salarial del INE, consiste en series de resultados promedio del coste laboral y salarial por trabajador. Tales series recogen tanto la variación del coste como los cambios en la composición del empleo.

2. Características del proyecto

2.1. Fuentes de información utilizadas

Como ya se ha indicado, la construcción del índice se realiza utilizando encuestas elaboradas por el INE: las Encuestas de Estructura Salarial Anual (EAES) y la Cuatrienal (EES). Estas encuestas han sido la fuente de información elegida para la elaboración del IPT por las siguientes razones fundamentales:

- ✓ Son una fuente de información ya existente.
- ✓ Su riqueza informativa: Disponen de datos de asalariados individuales, en especial de la variable ocupación, que es la central en este tipo de análisis.
- ✓ Su tamaño muestral permite el estudio de las características de los asalariados de forma muy detallada.
- ✓ Aunque no disponen de información de todos los componentes del coste laboral, sí la ofrecen sobre el más importante: el salario. También disponen de las horas normales de trabajo necesarias en el cálculo del Índice.

Estas encuestas recogen los salarios de forma individual y, junto a ellos, una gran cantidad de variables relacionadas con el asalariado. Gracias a esto es posible establecer relaciones entre el salario y algunas variables que pueden contribuir a determinar su cuantía.

Dichas variables son importantes a la hora de construir el IPT, pues debe medir los cambios en el precio de una "cesta básica" estandarizada de puestos de trabajo. La cuestión de determinar las características de los puestos de trabajo en el mercado laboral está estrechamente ligada al problema de definir los factores que determinan el precio del trabajo. La ocupación, es decir, las tareas y funciones que se realizan en el puesto de trabajo, es la variable central para la determinación de dicho precio; a esta hay que añadir otras variables, como las que caracterizan a la unidad en la que se desempeña el puesto de trabajo (actividad económica, tamaño, región, etc.), las características del contrato de trabajo (si es a tiempo completo o parcial, de duración determinada o indefinido), etc.. Todas estas variables existen en la Encuestas de Estructura Salarial.

2.2. Ámbito poblacional

Comprende a los asalariados por cuenta ajena que cotizan a la Seguridad Social bajo las siguientes condiciones:

- Pertenecen al Régimen General de la Seguridad Social y tienen actividad económica encuadrada en las Secciones B a S de la CNAE-09
- Pertenecen al Régimen Especial de Trabajadores del Mar y su actividad económica es el transporte marítimo (división 50 de la CNAE-09).
- Han estado de alta en la Seguridad Social durante al menos 2 meses en el año de referencia, siendo uno de ellos el mes de octubre.

Se excluye la Agricultura, los Hogares que emplean personal doméstico, las Organizaciones extraterritoriales y los funcionarios adscritos a Mutualidades.

Se excluyen los presidentes, miembros de consejos de administración y en, general, todo aquel personal cuya remuneración no sea principalmente en forma de salario, sino por comisiones o beneficios.

2.3. **Ámbito geográfico**

Se investiga todo el territorio nacional.

2.4. **Periodo de referencia**

El periodo de referencia para la información proporcionada por el IPT es anual.

2.5. **Unidades de análisis**

Teniendo en cuenta que el objetivo del IPT es medir el cambio en el precio de la mano de obra en el tiempo, la unidad de análisis es la ganancia por hora de trabajo de los asalariados.

2.6. **Variables de clasificación**

Las variables de clasificación que se utilizan para determinar los puestos de trabajo considerados en la elaboración del IPT son las siguientes:

- **Actividad:** las secciones de actividad según la CNAE-09 del ámbito muestral de las Encuestas de Estructura Salarial: B, C, D, E, F, G, H, I, J, K, L, M, N, O, P, Q, R y S.
- **Tamaño del centro de trabajo (estrato):** toma 8 valores distintos, según el número de asalariados del centro:
 - 1: de 1 a 4 asalariados
 - 2: de 5 a 9 asalariados
 - 3: de 10 a 19 asalariados
 - 4: de 20 a 49 asalariados
 - 5: de 50 a 99 asalariados
 - 6: de 100 a 199 asalariados
 - 7: de 200 a 499 asalariados
 - 8: 500 y más asalariados
- **Región:** Comunidad Autónoma en la que se encuentra el centro de trabajo:
 - 01: Andalucía (incluye 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)

- **Sexo:** mujeres (M) y varones (V).
- **Ocupación:** los grandes grupos de la Clasificación Nacional de Ocupaciones (CNO-11): del 1 al 9. Los grupos 6 y 7 aparecen unidos por problemas de representación muestral en la Encuestas de Estructura Salarial.
- **Tipo de contrato:** indefinido (1) y de duración determinada (2).
- **Edad del asalariado:** toma 5 valores distintos, determinados por los siguientes intervalos de edad:

- 01: menos de 25 años
- 02: de 25 a 34 años
- 03: de 35 a 44 años
- 04: de 45 a 54 años
- 05: 55 o más años

- **Nacionalidad del asalariado:** española (1) y resto de nacionalidades (2).
- **Antigüedad del asalariado en el centro de trabajo:** toma 6 valores distintos, determinados por los siguientes intervalos:

- 1: menos de 1 año de antigüedad
- 2: de 1 a 3 años de antigüedad
- 3: de 4 a 10 años de antigüedad
- 4: de 11 a 20 años de antigüedad
- 5: de 21 a 29 años de antigüedad
- 6: 30 o más años de antigüedad

3. Metodología de cálculo

3.1. Fórmula de cálculo del índice

Para la elaboración del IPT se ha utilizado un índice de Laspeyres encadenado.

La elección de un índice de Laspeyres encadenado se debe a que pueden existir cambios en la estructura del empleo de un año a otro y este tipo de indicador permite introducirlos anualmente a través de las ponderaciones sin tener que esperar a un cambio de base como ocurre en un índice de base fija.

La formulación del IPT se detalla a continuación:

- 1) Se denotan con c los cruces de las variables que definen las características de los puestos de trabajo (detalladas en el punto 2.6.). Es decir, se considera una estructura fija que es la constituida por los mencionados cruces, con el objetivo de ver cómo ha variado la ganancia por hora descontando el efecto composición.

G_c^k : ganancia por hora de los asalariados en el cruce c en el año k .

h_c^k : horas trabajadas por los asalariados en el cruce c en el año k .

- 2) Definición de la fórmula básica de Laspeyres que debe utilizarse para calcular el IPT durante el año j , para el año base k :

$$IPT_{j(k)} = \frac{\sum_c G_c^j h_c^k}{\sum_c G_c^k h_c^k}$$

- 3) Definición del coeficiente de enlace del año j con el año $j+1$:

$$L_{j,j+1} = \frac{\sum_c G_c^{j+1} h_c^j}{\sum_c G_c^j h_c^j}$$

- 4) Entonces la fórmula del índice de Laspeyres encadenada para el año j con referencia al año k es:

$$IPT_{j(k)} = 100 * L_{k,k+1} * L_{k+1,k+2} * \dots * L_{j-2,j-1} * IPT_{j(j-1)}$$

- 5) En este caso el año base será $k=2008$, por lo que su índice será 100. El coeficiente de enlace $L_{j,j+1}$ puede reescribirse como:

$$L_{j,j+1} = \frac{\sum_c G_c^{j+1} h_c^j}{\sum_c G_c^j h_c^j}$$

Es decir, el coeficiente de enlace se obtiene como el índice de un año en base al año anterior.

Por lo tanto, el índice de Laspeyres quedaría como sigue:

$$IPT_{j(k)} = 100 * IPT_{k+1(k)} * IPT_{k+2(k+1)} * \dots * IPT_{j-1(j-2)} * IPT_{j(j-1)} \text{ donde } k=2008.$$

6) La fórmula para calcular el índice encadenado de una combinación Q de cruces c en el año k sería la siguiente:

$$IPT_{j(k),Q} = 100 * IPT_{k+1(k),Q} * IPT_{k+2(k+1),Q} * \dots * IPT_{j-1(j-2),Q} * IPT_{j(j-1),Q}$$

3.2. Periodo base

El año 2008 es el periodo base, por lo que todos los índices que se calculen estarán referidos a ese año.

3.3. Periodo de referencia de las ponderaciones

La estructura de las ponderaciones establece la importancia de cada puesto de trabajo respecto al resto de ellos basado en el coste salarial realizado por el empleador. Para cada año, la medición de las ponderaciones se realiza a partir de las ganancias por hora y las horas trabajadas por los trabajadores asociados a cada tipo de puesto de trabajo en el año anterior.

3.4. Cálculo de la ganancia hora

Para obtener una estimación más robusta de la ganancia hora en los distintos puestos de trabajo definidos por las variables de clasificación, se utiliza un modelo de regresión hedónica. No es posible incorporar directamente datos observados, dado que hay puestos de trabajo que aparecen/desaparecen de un año a otro y que la representación muestral de algunos puestos es escasa.

El modelo de regresión utilizado en el cálculo de la ganancia hora es el siguiente:

Para cada año a, se asume que la ganancia por hora, G, para un asalariado que pertenece a un cruce c de las variables explicativas del modelo, es:

$$l_c^a = \ln G_c^a = x_c' \beta^a + \varepsilon_c^a \quad (1)$$

donde:

x_c' es un vector de dimensión (1xp) cuyos elementos son iguales a 0 o 1 en función de las características que definen al cruce c, en cuanto a los efectos principales y a las interacciones se refiere,

β^a es un vector de dimensión (px1) que contiene los parámetros desconocidos del modelo,

ε_c^a es la componente aleatoria del modelo para la celda c en el año a.

El vector β^a define el efecto proporcional sobre la ganancia por hora de las p variables dicotómicas incluidas en X_c^a . Los p parámetros desconocidos incluyen la constante y los coeficientes de las variables dicotómicas asociadas a los efectos principales y a las interacciones del modelo de regresión.

Para cada r categorías posibles que tiene un efecto principal, el modelo incluye (r-1) parámetros y si la interacción tiene (r x s) posibles combinaciones de valores, el modelo incluiría (r-1) x (s-1) parámetros.

Las componentes aleatorias verifican:

$$E[\varepsilon_c^a] = 0, \text{Var} [\varepsilon_c^a] = \sigma_a^2, \text{Cov} [\varepsilon_c^a, \varepsilon_d^a] = 0, \forall c \neq d \quad (2)$$

Una vez definido el modelo, se debe estimar el vector de parámetros β^a con la información disponible cada año. Para ello se formula el modelo (1) en notación matricial de la siguiente manera:

$$L^a = X^a \beta^a + \varepsilon^a \quad (3)$$

Donde

L^a es un vector de dimensión ($n^a \times 1$) que contiene los n^a elementos l_c^a del año a. Es decir, contiene tantas filas como asalariados tenemos en la muestra del año a.

X^a es una matriz de dimensión ($n^a \times p$), cuyos elementos son iguales a 0 o a 1. Cada fila representa a un asalariado y cada columna contiene las p características que definen a dicho asalariado.

β^a es un vector de dimensión ($p \times 1$) que contiene los parámetros desconocidos del año a. Incluye la constante y los coeficientes asociados a los efectos principales y las interacciones del modelo.

ε^a es un vector de dimensión ($n^a \times 1$) que contiene las n^a componentes aleatorias del modelo en el año a. Este vector verifica:

$$E[\varepsilon^a] = 0, \text{Var}[\varepsilon^a] = \sigma_a^2 I_{n^a \times n^a} \quad (4)$$

Como los datos utilizados en el modelo de regresión proceden de muestras extraídas de una población total con un determinado diseño muestral, utilizamos el estimador de mínimos cuadrados ponderados por los pesos de muestreo (MCOW) de β^a :

$$\hat{\beta}^a = (X^{a'} W^a X^a)^{-1} X^a W^a L^a \quad (5)$$

donde W^a es una matriz diagonal de dimensión $n^a \times n^a$ que contiene los pesos de muestreo.

Y su varianza es:

$$Var[\hat{\beta}^a] = \sigma_a^2 (X'^a W^a X^a)^{-1} X'^a W^{a2} X^a (X'^a W^a X^a)^{-1} = V^a \quad (6)$$

donde la matriz V^a tiene dimensión $(p \times p)$.

La ganancia por hora estimada correspondiente a cada cruce se obtiene a partir de la fórmula (1), con lo que la ganancia estimada en el cruce c el año a es la siguiente:

$$\hat{G}_c^a = \exp(x_c' \hat{\beta}^a) \quad (7)$$

El problema de este estimador es que tiene sesgo, pues aunque el estimador $\hat{\beta}^a$ sí es insesgado, al calcular su exponencial esta característica se pierde. Para corregir este sesgo utilizamos una variación del estimador propuesto por Shaarawi y Viveros (1997):

$$\hat{G}_c^a = \exp \left\{ x_c' \hat{\beta}^a - \frac{1}{2} x_c' \hat{V}^a x_c + \frac{1}{2} \hat{\phi}_a^2 \hat{\sigma}_a^2 \right\} \quad (8)$$

El estimador (8) corrige sustancialmente el sesgo del estimador (7), asumiendo la normalidad de los errores ε_c^a .

Para obtener la estimación de la varianza que aparece en las expresiones anteriores, se definen los residuos e_c^a como la diferencia entre los logaritmos de la ganancia por hora observada y la estimada:

$$e_c^a = l_c^a - x_c' \hat{\beta}^a \quad (9)$$

La varianza $\hat{\sigma}_a^2$ se estima con la media de cuadrados residual:

$$\hat{\sigma}_a^2 = \frac{1}{n^a - p} \sum_c^{n^a} (e_c^a)^2 \quad (10)$$

Siendo la estimación de $\hat{\phi}_a^2$:

$$\hat{\phi}_a^2 = 1 - \frac{\hat{\sigma}_a^2}{2(n^a - p)} - \frac{\hat{\sigma}_a^4}{3(n^a - p)^2} \quad (11)$$

No obstante, en (11) las varianzas del numerador son menores que 1 y al dividir las por el tamaño muestral resulta ser $\hat{\phi}_a^2 \cong 1$, por lo que la expresión (8) queda finalmente:

$$\hat{G}_c^a = \exp \left\{ x_c' \hat{\beta}^a - \frac{1}{2} x_c' \hat{V}^a x_c + \frac{1}{2} \hat{\sigma}_a^2 \right\} \quad (12)$$

4. Plan de difusión y periodicidad

4.1 PLAN DE DIFUSIÓN

Los datos se presentarán en formato de tablas, desagregados por las siguientes variables de clasificación:

- Sección de actividad
- Estrato
- Región
- Sexo
- Ocupación
- Tipo de contrato
- Edad
- Nacionalidad
- Antigüedad

También se facilitarán las ponderaciones utilizadas en cada año.

4.2 PERIODICIDAD DE LA DIFUSIÓN

Los resultados de esta operación se difundirán **anualmente**, previsiblemente en el último trimestre del año.

5. Calendario de implantación

Se iniciará con la difusión de los resultados correspondientes al año 2014, que tendrá lugar tras la publicación de los datos de la Encuesta Cuatrienal de Estructura Salarial de ese mismo año. Irán acompañados de una serie retrospectiva desde 2008.

6. Presupuesto estimado

Los recursos humanos asignados al proyecto son los siguientes:

Jefa de área

Director de programa

Jefe de servicio/sección

Estas personas no se dedicarán en exclusiva a la realización de este proyecto.

El presupuesto específico estimado para la operación en el periodo 2015-2018 (afecta a los Planes Estadísticos Nacionales 2013-2016 y 2017-2020) se cifra en 72,0 miles de euros, repartido por años y capítulos de la siguiente manera:

	CAP.1	CAP.2	CAP.4	CAP.6	TOTAL
2015	14,5				14,5
2016	17,4				17,4
2017	19,1				19,1
2018	21,0				21,0
					72,0