
Estadística Oficial

The regression model in the House Price Index

Alicia Gila García and María Novás Filgueira

Subdirección General de Estadísticas Coyunturales y de Precios
Instituto Nacional de Estadística

✉ alicia.gila.garcia@ine.es, maria.novas.filgueira@ine.es

Abstract

The Spanish House Price Index (HPI) is a quarterly indicator that measures the changes in the transaction prices of dwellings that households purchase. Indices are calculated according to a harmonised approach in order to give comparable measures across EU countries.

Location and characteristics of the purchased dwellings vary in each period, therefore prices need to be adjusted for quality changes. For that purpose HPI is compiled using a combination of stratification and hedonics. The regression model enables more detailed strata and consequently an improvement on quality change adjustments. This paper provides a technical description of the underlying model and how it works in the index calculation.

Keywords: IPC, IPCA, IPV, HPI.

AMS Subject classifications: 91B82

1. Introducción

El Índice de Precios de Vivienda (IPV) es un indicador trimestral, cuyo objetivo es medir la evolución del precio de compraventa de vivienda en España. Su origen se remonta al año 2002, cuando la Oficina de Estadística Europea (Eurostat) creó un grupo de Estudio Piloto sobre vivienda en propiedad, para medir la evolución de precios de la vivienda, englobando a su vez tanto los costes de adquisición como de mantenimiento de la misma. España ha formado parte de este grupo desde su inicio, compuesto en la actualidad por la mayoría de los países de la UE.

Los trabajos desarrollados en el seno de este grupo culminaron con la publicación del IPV en 2008. En este sentido, el IPV se ha concebido bajo los mismos criterios que el Índice de Precios de Consumo Armonizado (IPCA) y por tanto, permite establecer comparaciones en la evolución de precios de la vivienda entre los distintos Estados Miembros de la UE.

Un aspecto crucial en cualquier índice de precios es el tratamiento de los cambios de calidad de los bienes cuyos precios se siguen en el tiempo. El IPV hace especial hincapié en el ajuste por los cambios que se producen en la composición y calidad de las viviendas que forman parte del cálculo del índice, de unos trimestres a otros. Para ello, emplea un modelo econométrico, lo que confiere a esta estadística un carácter diferenciador frente a las que se elaboran desde el sector público hasta ahora. Este documento proporciona una descripción de la fuente de información seleccionada y el tratamiento de los datos recibidos. A continuación se explica en detalle el modelo de regresión y de como interviene en el proceso de cálculo del índice. Para los detalles de la estimación de los modelos, se puede consultar Draper (1998), Montgomery (2001) y Peña (1993 y 2002).

Por último, se muestra la evolución de precios de la vivienda desde el comienzo de la serie en 2007, para el total de viviendas y por separado para viviendas nuevas y de segunda mano.

2. Fuente de información

La fuente de información seleccionada para el cálculo del IPV es el registro de notarios, que contiene entre otros datos, los precios oficiales de todas las compraventas que han tenido lugar en el territorio español. Se trata de la mejor fuente teniendo en cuenta criterios como la disponibilidad, puntualidad, cobertura y contenido de la información.

De acuerdo con Eurostat, los precios de las compraventas de viviendas realizadas en el trimestre son los que deben formar parte del cálculo del índice, no se permite por tanto la utilización de precios de oferta o valores de tasación. Asimismo, se incluye todo tipo de viviendas, tanto nuevas como de segunda mano, pisos o viviendas unifamiliares, e independientemente del modo de pago.

En cuanto a la cobertura poblacional y geográfica del IPV, se recogen todas las transacciones de viviendas adquiridas por hogares particulares, residentes y no residentes, dentro del territorio económico. Así, quedan fuera del ámbito de este indicador las viviendas adquiridas por personas jurídicas o entidades financieras. Del mismo modo, siguiendo criterios del IPCA, se excluyen las viviendas de protección oficial por ser de acceso restringido.

El Consejo General del Notariado a través de la Agencia de Certificación Notarial (ANCERT) envía periódicamente al INE información sobre las transmisiones de viviendas realizadas. Aproximadamente 45 días finalizado el trimestre de referencia de los datos, se recibe el fichero de transacciones que se utiliza en el cálculo del IPV, éste contiene más del 90 % de las compraventas de viviendas realizadas en el trimestre.

La base de datos de los notarios, además del precio de transmisión de la

vivienda, contiene información detallada sobre la calidad de la misma, su localización exacta, tamaño, tipo, etc. Asimismo, incluye datos adicionales que permiten la adaptación al ámbito de estudio, como el tipo de comprador y si la vivienda es de precio libre o de protección oficial.

3. Base metodológica

Uno de los aspectos más relevantes en cualquier índice de precios es el ajuste por el cambio que se produce en la calidad de los bienes cuyos precios se siguen a lo largo del tiempo. Sin embargo, mientras que para cualquier artículo de la cesta de la compra del IPC/IPCA, puede ser relativamente sencillo seguir su precio periódicamente, manteniendo fijas sus características, esto resulta imposible cuando se trata de viviendas. De hecho, la localización y características de las viviendas utilizadas en el cálculo del índice es diferente cada trimestre. Por ello, para una correcta medición de la evolución de precios, es necesario ajustar los cambios en la composición y en la calidad de las viviendas que se adquieren de unos periodos a otros.

Una posible solución puede ser la creación de estratos o tipologías de viviendas que agrupen aquellas con características similares. Los precios medios de cada estrato se pueden observar trimestralmente y obtener la evolución conjunta mediante una media ponderada. En este sentido, conviene crear estratos pequeños con viviendas muy similares entre sí; no obstante, para que estos precios medios sean representativos es necesario un número mínimo de observaciones por estrato y trimestre. Esto puede hacer que los estratos no sean tan detallados como sería deseable. Por ello, el IPV combina estratificación con hedónica, esto permite definir los estratos de una forma más precisa al no requerir un número mínimo de viviendas por estrato, y así el ajuste de calidad mejora considerablemente.

Los modelos hedónicos se usan habitualmente en el cálculo de índices de precios para controlar las variaciones de calidad, en los productos que integran los índices. Estos modelos pretenden explicar el valor de un bien en función de cada uno de los atributos o características que lo componen. Con ello es posible determinar cómo cambiará dicho valor al variar la cantidad en que se encuentra presente cada uno de estos atributos, y consecuentemente, predecir precios.

Los precios estimados por el modelo de regresión se combinan con las ponderaciones de cada estrato o tipología de vivienda para obtener el IPV. Mientras que el conjunto de tipologías y sus ponderaciones permanecen fijas durante un año, los precios se estiman trimestralmente; de este modo se obtienen variaciones puras de precios.

En primer lugar se desarrolla el elemento más importante de esta metodología, el modelo de regresión (sección 4). Posteriormente se muestra como se

calculan los dos componentes fundamentales del índice: ponderaciones (sección 5) y precios (sección 6).

4. Características del modelo de regresión

El modelo de regresión que emplea el IPV tiene como variable dependiente el logaritmo neperiano del precio por metro cuadrado de la vivienda. La relación no lineal entre el precio y la superficie de la vivienda, dado que esta última entra en el modelo también como variable independiente, hace que el modelo semilogarítmico sea adecuado.

Anualmente se analizan las variables independientes que conforman el modelo de regresión, vigente durante todo un año. Para ello se utiliza la información que contienen las observaciones que han participado en el cálculo del índice durante el año anterior. Una variable será incluida en el modelo, si es significativa y sus coeficientes son coherentes durante todos los trimestres estudiados.

Las once variables explicativas del modelo son todas categóricas (toman un número finito de valores) y se pueden clasificar en dos grupos, uno relativo a las características físicas de la vivienda y otro a su localización.

- Características de la vivienda:

- Nueva/segunda mano
- Tipo de vivienda (piso/unifamiliar)
- Garaje (si/no)
- Trastero (si/no)
- Cooperativa (si/no)
- Superficie (10 intervalos)

La superficie es una variable numérica cuyos valores se agrupan en 10 intervalos de distinta amplitud, teniendo en cuenta la relación entre precio y superficie (los intervalos de mayor amplitud incluyen viviendas de mayor tamaño):

< 40 ; $[40, 60)$; $[60, 75)$; $[75, 90)$; $[90, 105)$; $[105, 120)$;
 $[120, 150)$; $[150, 180)$; $[180, 240)$ y $\geq 240 m^2$

- Planta (seis categorías)

Los pisos se clasifican en seis categorías: sótano, bajo, primero, segundo, resto de plantas y ático.

- Localización de la vivienda

- Cluster de provincias (6 grupos)

Clasificación de las 52 provincias en 6 grupos, por medio de un análisis cluster, utilizando el importe medio de la hipoteca por provincia (INE).

- Tamaño del municipio (4 categorías)
Clasificación de los municipios en capital, no capital con más de 100.000 habitantes, municipios medianos (10.000-100.000) y pequeños (menos de 10.000); con datos de población del Padrón Municipal de Habitantes (INE).
- Turístico (4 tipos)
Se considera que un municipio es turístico cuando concentra un número elevado de pernoctaciones en alojamientos turísticos y/o un alto porcentaje de viviendas secundarias. Además, se tiene en cuenta su localización (si es costero o no) y su población para clasificarlo en alguna de las cuatro categorías siguientes:
 - * Turismo de sol y playa. Municipios turísticos costeros.
 - * Turismo cultural o de negocios. Capitales y otros municipios turísticos no costeros de más de 10.000 habitantes.
 - * Turismo rural o de naturaleza. Municipios turísticos de interior con menos de 10.000 habitantes.
 - * Resto de municipios.
- Tipo de entorno (14 tipos)
La clasificación de los códigos postales en 14 categorías se hace con información del Censo de Población y Viviendas de 2001 sobre la condición socio-económica media y el nivel medio de problemas. Estos problemas son la delincuencia, el nivel de ruido o suciedad en la zona, las malas comunicaciones, la escasez de zonas verdes, etc. Asimismo, se tiene en cuenta el precio medio del metro cuadrado por código postal obtenido con los datos de los notarios del año anterior.

Estas variables de localización se actualizan anualmente con la última información disponible de las fuentes empleadas en cada caso.

Además, el modelo incluye las interacciones dobles más significativas entre estos efectos principales. Una interacción es el producto de dos efectos principales; por ejemplo, nueva/segunda mano \times garaje es significativa, si la variación de precio de la vivienda, por el hecho de incluir garaje, es distinto si se trata de una vivienda nueva a si es de segunda mano (en general el incremento de precio es menor en las nuevas).

Las interacciones del modelo se revisan cada año, para ello se realizan pruebas con las observaciones que se han utilizado en el cálculo del índice durante el año anterior. Los criterios seguidos para la selección de las interacciones son tres:

- Debe ser significativa.
- Su inclusión debe contribuir a la capacidad explicativa del modelo tanto como sea posible.

- El número de observaciones trimestrales deberá ser superior a 50 en cada uno de los pares de valores posibles de la interacción.

Cada interacción añade restricciones al modelo, concretamente precio observado y estimado para cada cruce de valores de la interacción debe coincidir, por ello se exige ese mínimo de observaciones del tercer criterio.

Tanto los efectos principales como las interacciones entran en el modelo a través de variables dicotómicas. Por cada r valores diferentes que toma una variable, el modelo introduce $(r - 1)$ variables dicotómicas (una de las categorías se omite) y estima $(r - 1)$ parámetros. Estos parámetros representan los precios implícitos de las características. De manera similar, si una interacción toma $r \times s$ pares de valores, entonces se estiman $(r - 1) \times (s - 1)$ parámetros. Por ejemplo, en 2012, el modelo incluye 135 parámetros. Aquí radica la principal ventaja del uso de un modelo de regresión, porque con poco más de 100 parámetros es posible estimar precios para más de 66.000 tipologías de vivienda.

5. Tipologías de vivienda y sus ponderaciones

Las variables que forman parte del modelo de regresión son las que determinan el conjunto de tipologías, cuyos precios se van a seguir trimestre a trimestre.

De este modo, cada combinación de valores que toman las once variables del modelo, presentes en alguna de las transmisiones de los tres últimos años, determina una celda o tipología de vivienda.

En 2012 el número de tipologías que conforman el índice supera las 66.000. La ponderación de cada tipología de vivienda representa la importancia que esta tiene frente al total de compra de vivienda. Cada año a , se calcula la ponderación para cada celda c del siguiente modo:

$$W_c^a = \frac{Q_c^{(a-1, a-2, a-3)} \times \hat{P}_c^{4, a-1}}{\sum_c Q_c^{(a-1, a-2, a-3)} \times \hat{P}_c^{4, a-1}}$$

donde tanto las cantidades como los precios se refieren a la misma unidad, el metro cuadrado de vivienda. Por ello, $Q_c^{(a-1, a-2, a-3)}$ representa la cantidad anual media de metros cuadrados de las viviendas vendidas en el periodo de referencia de las ponderaciones (a-1, a-2, a-3), pertenecientes a la celda c ; y $\hat{P}_c^{4, a-1}$ el precio estimado de la celda c en el cuarto trimestre del año anterior. De esta manera, la ponderación anual de cada tipología de vivienda representa el gasto efectuado en la compra de viviendas de ese tipo (frente al total de viviendas) realizada en los tres años anteriores y valoradas a precios del cuarto trimestre del año anterior.

Por una parte, el hecho de utilizar tres años completos en las cantidades hace que la estructura de ponderaciones no fluctúe en exceso año a año, evitando

así que los cambios en las cantidades tengan efecto en el cálculo de variaciones anuales de precios. Asimismo, se consigue un mayor número de tipologías de viviendas que empleando un único año.

Por otra parte, los precios que intervienen en el cálculo de las ponderaciones son los precios estimados del cuarto trimestre del año anterior, los mismos con los que se calculan los índices elementales en la fórmula de Laspeyres, empleada en el IPV, que se verá en el apartado de índices.

6. Precios

Los precios se estiman trimestralmente aplicando el modelo de regresión sobre el fichero con las compraventas realizadas en el trimestre. Previamente, los datos recibidos pasan por un proceso de validación que se describe a continuación.

6.1. Validación de los datos

La localización de la vivienda es un aspecto esencial en la explicación del precio de la misma, por ello, del proceso de depuración trimestral cabe destacar la comprobación de la combinación de las tres variables geográficas: provincia, municipio y código postal en cada observación. En caso de error, se intenta subsanar de forma manual.

Además, para que una observación intervenga en la estimación de parámetros del modelo de regresión es necesario que tenga todas las variables del modelo informadas. Una característica relevante de los datos recibidos de los notarios es la completitud de las observaciones. No obstante, en algunos casos es necesaria la imputación en las variables superficie y tipo de entorno.

Para imputar el valor perdido de la variable superficie en una observación, se calculan diez posibles precios del metro cuadrado, dividiendo el precio de la vivienda entre las marcas de clase de los diez intervalos de superficie vistos en la sección 4.

Por otro lado, se calcula el precio medio del metro cuadrado del código postal al que pertenece dicha observación en ese trimestre (como media geométrica de los precios del metro cuadrado de las viviendas de la muestra trimestral que pertenecen a ese código postal). Por último, se asigna como valor de la superficie de la observación incompleta, la marca de clase del intervalo cuyo precio del metro cuadrado sea más parecido al precio medio del metro cuadrado del código postal al que pertenece la observación incompleta en dicho trimestre.

De forma similar, con cada observación sin valor en la variable tipo de entorno se compara el precio del metro cuadrado de dicha observación con los precios medios del metro cuadrado de todos los códigos postales pertenecientes a su mismo municipio (estos precios medios se calculan como se ha mencionado

anteriormente). Se asigna, en cada caso, el valor de la variable tipo de entorno del código postal con el precio medio del metro cuadrado más parecido al precio del metro cuadrado de la observación.

Una vez depurado y completado el fichero de observaciones del trimestre, se utiliza el modelo de regresión para estimar los precios.

6.2. Especificación del modelo

En el trimestre q , se asume que el precio del metro cuadrado, P , de la vivienda i perteneciente a la celda o tipología de vivienda c , es:

$$l_{i,c}^q = \ln P_{i,c}^q = x'_c \beta^q + \varepsilon_{i,c}^q \quad (6.1)$$

donde:

x'_c es un vector de dimensión $(1 \times p)$ cuyos elementos son iguales a 0 o 1 en función de las características que definen a la celda c , en cuanto a efectos principales e interacciones se refiere,

β^q es un vector de p parámetros desconocidos, de dimensión $(p \times 1)$, y

$\varepsilon_{i,c}^q$ es la componente aleatoria del modelo, en el trimestre q .

El vector β^q define el efecto proporcional sobre el precio esperado del metro cuadrado de vivienda de las p variables dicotómicas incluidas en x'_c , en el trimestre q . Los p parámetros desconocidos incluyen la constante y los parámetros de las variables dicotómicas asociados a los efectos principales y a las interacciones del modelo.

Las perturbaciones $\varepsilon_{i,c}^q$ verifican:

$$E(\varepsilon_{i,c}^q) = 0, \text{Var}(\varepsilon_{i,c}^q) = \sigma_q^2, \text{Cov}(\varepsilon_{i,c}^q, \varepsilon_{j,d}^q) = 0 \quad \forall (q, i, c) \neq (q', j, d)$$

Una vez definido el modelo que estará en vigor durante un año, el vector β^q se estima cada trimestre con la información disponible. Para ello se formula el modelo (6.1) en notación matricial, de la siguiente manera:

$$L^q = X^q \beta^q + \varepsilon^q$$

donde:

L^q es un vector de dimensión $(n^q \times 1)$ que contiene los n^q elementos $l_{i,c}^q$ del trimestre q . Es decir, contiene tantas filas como compraventas de viviendas se han realizado en el trimestre (n^q) ,

X^q es una matriz de dimensión $(n^q \times p)$, cuyos elementos son iguales a 0 o 1. En esta matriz cada fila representa una vivienda y cada columna contiene una de las p características que definen dicha vivienda, en el trimestre q ,

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

$$E(\varepsilon^q) = 0, \text{Var}(\varepsilon^q) = \sigma_q^2 I_{n^q \times n^q}$$

El estimador MCO (mínimos cuadrados ordinarios) de β^q es:

$$\hat{\beta}^q = (X^{q'} X^q)^{-1} X^{q'} L^q$$

y su varianza es:

$$\text{Var}(\hat{\beta}^q) = \sigma^2 (X^{q'} X^q)^{-1} = V^q$$

El vector de parámetros $\hat{\beta}^q$ varía según los datos de cada trimestre y es el elemento fundamental utilizado para la estimación del precio medio por celda.

En la práctica, el IPV utiliza el estimador de mínimos cuadrados ponderados para corregir la heterocedasticidad, presente en las perturbaciones respecto a una de las variables del modelo y para asignar distintos pesos a las observaciones con valores imputados.

El modelo ponderado es el siguiente:

$$(W^q)^{1/2} L^q = (W^q)^{1/2} X^q \beta^q + (W^q)^{1/2} \varepsilon^q$$

donde W^q es una matriz diagonal de dimensión $(n_q \times n_q)$. Cada elemento de la diagonal principal corresponde a una vivienda y es el peso con el que entra esa vivienda en la regresión. Estos pesos se obtienen como producto de dos coeficientes, uno que corrige la heterocedasticidad y otro en función de los valores imputados de la observación.

Las nuevas expresiones para los estimadores $\hat{\beta}^q$ y su varianza son:

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

$$\text{Var}(\hat{\beta}^q) = \sigma^2 (X^{q'} W^q X^q)^{-1}$$

La varianza residual del modelo corregido se obtiene con los residuos ponderados:

$$e_{i,c}^q = \sqrt{w_i^q} (l_{i,c}^q - x_c' \hat{\beta}^q)$$

donde w_i^q es el elemento (i, i) de la matriz W^q .

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

$$\hat{\sigma}_q^2 = \frac{1}{n_q - p} \sum_{c,i}^{n^q} (e_{i,c}^q)^2$$

Finalmente, a partir de la fórmula (6.1), el precio estimado de la celda c , en el trimestre q es el siguiente:

$$\hat{P}_c^q = \exp(x'_c \hat{\beta}^q)$$

Aunque este estimador tiene una expresión sencilla, su sesgo es alto, para corregirlo se utiliza el estimador propuesto por El-Shaarawi y Viveros (1997):

$$\hat{P}_c^q = \exp\left\{x'_c \hat{\beta}^q - \frac{1}{2} x'_c \hat{V}^q x_c + \frac{1}{2} \hat{\varphi}_q^2 \hat{\sigma}_q^2\right\} \quad (6.2)$$

donde

$$\hat{\varphi}_q^2 = 1 - \frac{\hat{\sigma}_q^2}{2(n_q - p)} - \frac{\hat{\sigma}_q^4}{3(n_q - p)^2}.$$

6.3. Aplicación del modelo de regresión para la estimación de precios

Toda la formulación anterior se resume, en la práctica, en la realización de siete modelos lineales generales en SAS.

Inicialmente se realiza un modelo lineal general sobre todas las observaciones resultantes tras el proceso de validación. A partir de los residuos de este primer modelo se determinan y eliminan las observaciones atípicas.

Posteriormente, como la variabilidad de los residuos en las observaciones donde se ha imputado el valor de alguna de las variables explicativas del modelo, es mayor que en el conjunto de aquellas que vienen completas en el fichero de datos, a las observaciones completas se les asigna en la regresión un peso igual a uno, mientras que aquellas con valores imputados, tienen asignado un peso inferior. A continuación se calcularán las ponderaciones con las que entran las observaciones con valor imputado en la superficie, en el tipo de entorno ó en ambas. Para ello se seleccionan las observaciones completas y se realiza el modelo lineal general con los 11 efectos principales y las 9 interacciones, se obtiene así el Error Cuadrático Medio (MCE_T). A su vez, se aplican tres modelos más, en el primer caso se elimina la superficie como efecto principal y las interacciones en las que aparezca, para obtener MCE_{T-Sup} . Se procede del mismo modo con el tipo de entorno ($MCE_{T-Entorno}$) y el último modelo se realiza eliminando ambas variables a la vez y las interacciones en las que aparezcan ($MCE_{T-Sup-Entorno}$).

Los coeficientes se definen del siguiente modo:

$$\lambda_i = \begin{cases} 1 & \text{observación } i \text{ sin valores imputados} \\ \frac{MCE_T}{MCE_{T-Sup}} & \text{observación } i \text{ con superficie imputada} \\ \frac{MCE_T}{MCE_{T-Entorno}} & \text{observación } i \text{ con tipo de entorno imputado} \\ \frac{MCE_T}{MCE_{T-Sup-Entorno}} & \text{observación } i \text{ con ambas variables imputadas} \end{cases}$$

Como el modelo completo tiene una varianza residual inferior a la del sub-modelo que excluye uno o varios efectos principales (y sus correspondientes

interacciones), se verifica que: $0 \leq \lambda_i \leq 1$.

Los residuos del modelo ponderado anterior presentan signos de heterocedasticidad respecto a la variable tipo de entorno. Por ello, se corrige la heterocedasticidad calculando otros pesos a través de los residuos estandarizados del modelo ponderado.

Por último, para recoger el efecto de las dos ponderaciones, se realiza un modelo lineal general ponderado, con ponderación igual al producto de las dos anteriores. Como resultado se obtiene la estimación del vector de parámetros β , donde cada componente representa el porcentaje en el que se verá incrementado el precio, al poseer la característica correspondiente. Estos parámetros estimados permiten calcular el precio de cada tipología de vivienda siguiendo la fórmula (6.2).

7. Índices

La fórmula general utilizada para el cálculo del IPV es un índice de Laspeyres encadenado, análoga a la utilizada en el IPC/IPCA. Los índices encadenados permiten cada año la revisión del modelo, del conjunto de tipologías de viviendas y sus ponderaciones, de manera que se trata de un indicador dinámico que se adapta de forma continua a la situación del mercado.

El índice en el trimestre q del año a referido al cuarto trimestre del año anterior, se obtiene como cociente de precios medios ponderados por unas cantidades, las mismas en numerador y denominador (correspondientes a los tres años previos) mientras que los precios corrientes (numerador) se comparan con los del cuarto trimestre del año anterior (denominador); ambos son precios estimados por el mismo modelo de regresión:

$${}_{(4,a-1)}I^{q,a} = \frac{\sum_c Q_c^{(a-1,a-2,a-3)} \times \hat{P}_c^{q,a}}{\sum_c Q_c^{(a-1,a-2,a-3)} \times \hat{P}_c^{q,a-1}}$$

La fórmula anterior se puede expresar también como media ponderada de índices elementales:

$${}_{(4,a-1)}I^{q,a} = \sum_c W_c^a \times {}_{(4,a-1)}I_c^{q,a}$$

donde:

$$W_c^a = \frac{Q_c^{(a-1,a-2,a-3)} \times \hat{P}_c^{q,a-1}}{\sum_c Q_c^{(a-1,a-2,a-3)} \times \hat{P}_c^{q,a-1}}$$

$${}_{(4,a-1)}I_c^{q,a} = \frac{\hat{P}_c^{q,a}}{\hat{P}_c^{q,a-1}} \times 100.$$

Como se puede observar, los precios que intervienen en el cálculo de las ponderaciones son los mismos con los que se comparan los precios del trimes-

tre corriente en los índices elementales. Por último, estos índices referidos al cuarto trimestre del año anterior se encadenan para obtener el índice en base 2007, índices publicados a partir de los cuales se calculan las tasas de variación:

$${}_{07}I^{q,a} = {}_{07}I^{4,a-1} \times \frac{({}_{4,a-1})I^{q,a}}{100}.$$

Para un mayor detalle acerca de la fórmula de cálculo del IPV se pueden consultar [3] y [4].

8. Evolución de precios

En la Figura 1 están representados los índices en base 2007, a nivel nacional para el total de viviendas y por separado para vivienda nueva y de segunda mano.

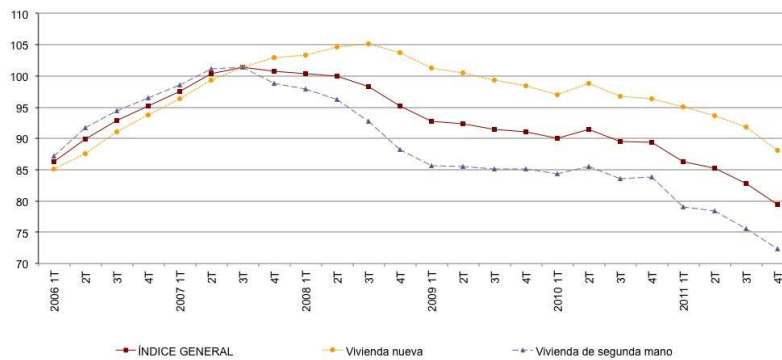


Figura 1: IPV, base 2007.

El nivel de precios actualmente es bastante inferior al del inicio de la serie, año 2006. La vivienda de segunda mano crece hasta el tercer trimestre de 2007 y a partir de este momento decrece severamente. Un comportamiento similar tiene la vivienda nueva aunque la bajada se produce un año después. En general, las bajadas de precios son muy acusadas en 2008 y a partir de 2010 mientras que en 2009 los descensos fueron moderados. También cabe destacar las fluctuaciones o picos que se muestran en el año 2010, debidas fundamentalmente a dos hechos: la subida del IVA del 7 al 8% a partir de julio de 2010 y la retirada de la desgravación fiscal por compra de vivienda a partir de enero de 2011. Ambas medidas provocaron un aumento considerable de las compraventas de viviendas en los trimestres previos a su entrada en vigor, según los datos del Ministerio de Fomento (40% en el segundo trimestre de 2010 y más del 80% en el cuarto trimestre), seguidos de una disminución (en torno al 50% en ambos casos) en los trimestres donde comenzaron a

ser efectivas. Algo similar a lo sucedido con las cantidades se mostró en los precios, aunque a niveles muy inferiores.

Desde 2011 el ritmo de decrecimiento es notable y prácticamente igual en ambos tipos de vivienda; sin embargo, mientras que la vivienda nueva está a niveles de precios de 2006, la de segunda mano está muy por debajo.

Finalmente, la caída de precios desde el nivel más alto de la serie (tercer trimestre de 2007) hasta el cuarto trimestre de 2011 es del 22%. Por tipo de vivienda, el descenso es del 16% en vivienda nueva y del 29% en la de segunda mano.

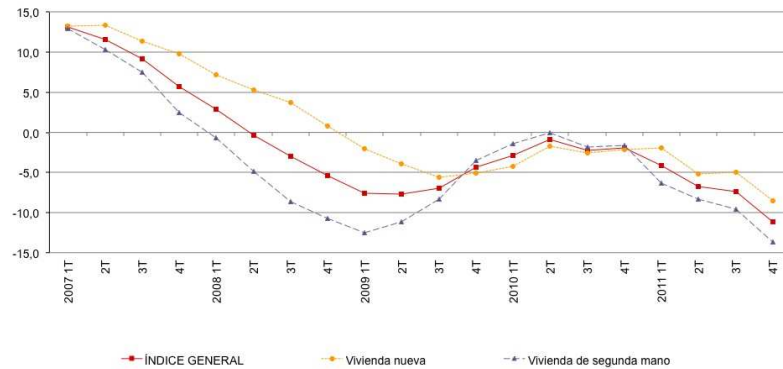


Figura 2: IPV. Tasas de variación anual.

En la Figura 2 aparecen representadas las tasas anuales con el mismo desglose que los índices. Cabe destacar el contraste de las tasas anuales al inicio de la serie, donde eran positivas y alcanzaban los dos dígitos y las tasas en el momento actual, con dos dígitos, pero en este caso negativas. Las tasas anuales de ambos tipos de viviendas decrecen desde el año 2007 aunque a un ritmo diferente. Las tasas de vivienda de segunda mano decrecen a un ritmo vertiginoso, alcanzando la primera tasa negativa en el primer trimestre de 2008, frente a las de vivienda nueva que consiguen el mismo efecto un año después.

Las variaciones anuales de precios de ambos tipos de vivienda han sido negativas desde entonces, aunque cabe destacar un pequeño repunte que abarca aproximadamente el año 2010, donde las tasas continuaron negativas pero no bajaron del -5% . A partir de 2011 comienzan de nuevo los descensos que vienen aconteciendo desde el inicio de la serie.

Referencias

- [1] Draper N.R. (1998). *Applied regression analysis*. John Wiley & Sons.

- [2] El-Shaarawi A.H. y Viveros R. (1997). *Inference about the mean in log–regression with environmental applications*. *Environmetrics*, 8, 569-582.
- [3] INE (2009). Índice de Precios de Vivienda. Metodología. En: <http://www.ine.es/daco/daco42/ipv/metodologia.pdf>
- [4] INE (2008). Housing Price Index (HPI), base 2007. Main characteristics. En: <http://www.ine.es/daco/daco42/ipv/notainfo.pdf>
- [5] INE (2004). Transacciones inmobiliarias (compraventas). En: <http://www.fomento.gob.es>
- [6] Montgomery D. C. (2001). *Introduction to linear regression analysis*. John Wiley & Sons.
- [8] Peña D. (1993). *Estadística, modelos y métodos. 2 Modelos lineales y series temporales*. Alianza Universidad Textos.
- [8] Peña D. (2002). *Regresión y diseño de experimentos*. Alianza Universidad Textos.

Acerca de los autores

Alicia Gila García es licenciada en Ciencias Matemáticas por la Universidad Complutense de Madrid (1996). Pertenece al Cuerpo de Diplomados en Estadística del Estado desde 1998 y al Cuerpo Superior de Estadísticos del Estado desde 1999. Durante los nueve primeros años en el INE ha ocupado distintos puestos dentro del Área de Metodología del Índice de Precios de Consumo (IPC). En la actualidad es jefa de Área en la S.G. de Estadísticas Coyunturales y de Precios, responsable del Índice de Precios de Vivienda (IPV). Entre sus funciones se encuentra la producción trimestral del IPV y la revisión anual de su metodología de cálculo.

María Novás Filgueira es licenciada en Ciencias Matemáticas por la Universidad de Oviedo (2000). Pertenece al Cuerpo de Diplomados en Estadística del Estado desde 2003 y al Cuerpo Superior de Estadísticos del Estado desde 2005. En la actualidad es Directora de Programa en el Área de Índice de Precios de Vivienda en la S.G. de Estadísticas Coyunturales y de Precios del INE. Entre sus funciones se encuentra el desarrollo metodológico del IPV, cumpliendo los criterios establecidos internacionalmente y la explotación de los datos para la producción trimestral del índice.