

Índice de Precios de Vivienda

Metodología

Madrid, 2009

Índice

1. Introducción	3
2. Antecedentes. Grupo de Estudio de Eurostat	5
3. Objetivos	8
4. Ámbitos de la investigación	9
5. Fuente de información. Variables utilizadas	10
6. Metodología de cálculo	13
6.1. Precios	14
6.2. Modelo de regresión	14
6.3. Ponderaciones	16
6.4. Cálculo de índices	18
6.5. Cálculo de tasas de variación	20
6.6. Cálculo de repercusiones	22
7. Difusión	26
Anexo I Glosario de términos	27
Anexo II Modelo de regresión	29
Especificación del modelo de regresión	29
Estimación de precios	30
Corrección de heterocedasticidad	31

1. Introducción

La presente metodología detalla las principales características y procedimientos utilizados en el cálculo del Índice de Precios de Vivienda (IPV). Se trata de un indicador que ha sido concebido con el objetivo de completar la información que se produce en España sobre el mercado de la vivienda.

Hasta la publicación del IPV, la principal estadística que proporcionaba información sobre la evolución de los precios del mercado inmobiliario era la producida por el Ministerio de la Vivienda. Por su parte, el sector privado, principalmente las compañías tasadoras y agencias inmobiliarias, publican también alguna información sobre precios de vivienda.

Cualquiera de ellas permite obtener una visión del mercado inmobiliario en España, pero no muestran la situación del sector desde todos los ámbitos posibles, lo cual impide realizar un análisis completo del mismo. Esta carencia de información es uno de los principales motivos que impulsaron al Instituto Nacional de Estadística (INE) a plantearse la realización de una estadística que cubriera el mercado de la vivienda de forma general, haciendo hincapié en los procedimientos de regresión hedónica para el tratamiento de los ajustes de calidad, aspecto de vital importancia en los índices de precios.

Esta misma inquietud está presente en los países de la Unión Europea (UE) desde hace años. Por ello, en el año 2002 se creó en el seno de la Oficina de Estadística Europea (Eurostat), un Grupo de Estudio para abordar el problema de la medición de los precios de la compra de vivienda. 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 realizados dentro del Grupo de Estudio se han desarrollado en tres fases, en cada una de ellas se han ampliado los objetivos perseguidos, que han pasado del estudio de la situación del mercado en cada uno de los países (fuentes de información disponibles y métodos de cálculo válidos) de la primera fase, a la producción periódica de índices que permitan medir la evolución de los precios de vivienda, tanto nueva como de segunda mano, objetivo de la tercera etapa, que finaliza a mediados del año 2009.

Con esta situación de partida, se hace imprescindible la implantación de una estadística destinada a conocer la evolución temporal de los precios de la compra de vivienda en España, que resuelva las necesidades de información actuales y permita establecer comparaciones internacionales.

Para ello, se diseñó un plan de trabajo con el objetivo de implantar un Índice de Precios de Vivienda (IPV), concebido teniendo en cuenta los trabajos realizados en el seno del Grupo de Estudio de Eurostat.

Por su parte, el INE creó en el año 2005 un Grupo de Trabajo Interministerial y del Banco de España, con representantes de Presidencia del Gobierno, Banco de España y los Ministerios de Economía y Hacienda, Justicia y Vivienda. En este Grupo de Trabajo se presentaron y analizaron todos los elementos que deben conformar un indicador de la evolución de los precios de la vivienda, con las características básicas que permitan la comparación internacional requerida.

El presente documento detalla el procedimiento seguido en los años previos a la implantación del IPV, así como las características metodológicas más relevantes que lo definen.

Se ha hecho especial hincapié en el método desarrollado para el tratamiento de la información obtenida, basado en modelos de regresión, lo que confiere a esta estadística un carácter diferenciador frente a las que se habían elaborado desde el sector público hasta ahora.

2. Antecedentes. Grupo de Estudio de Eurostat

La creación del Grupo de Estudio sobre la vivienda en propiedad coordinado por Eurostat sirvió para impulsar el desarrollo de una nueva operación destinada a medir la evolución de los precios de la vivienda en España. De hecho, los trabajos realizados dentro de este Grupo pueden ser considerados como los preliminares del proyecto de trabajo seguido para la obtención del IPV.

La **primera fase** del proyecto (entre los años 2002 y 2004) este Grupo, compuesto por Alemania, España, Finlandia, Polonia y el Reino Unido, tenía el cometido de estudiar la situación del mercado inmobiliario en cada uno de los países, localizar las principales fuentes de información disponibles y proponer un procedimiento estadístico de cálculo que permitiera realizar comparaciones de la evolución de los precios de la vivienda nueva entre los países de la UE.

Durante esta fase, en España se llevó a cabo una prueba piloto de recogida de precios de vivienda nueva, que tuvo como objetivos principales comprobar el funcionamiento de una recogida en campo y poner a prueba un modelo de cuestionario para, finalmente, analizar las diferencias existentes entre este sistema de recogida de información y la obtención de la misma a través de las compañías tasadoras.

Con este fin, la prueba en campo se circunscribió a las provincias de Madrid y Segovia, donde se visitaron las promociones de vivienda nueva para recoger el precio de venta y las características de todos los tipos de viviendas existentes en la promoción.

Asimismo, se realizaron estudios preliminares utilizando valores de tasación, para lo cual se trabajó con valores medios de las tasaciones realizadas por las más importantes compañías tasadoras de España en ocho provincias, distinguiendo por tamaño de la vivienda.

Las conclusiones obtenidas de estas primeras pruebas sirvieron para comenzar a trabajar sobre el diseño del futuro índice. A partir de ellas, se concibieron las ideas básicas para el desarrollo de la metodología del índice, tales como la periodicidad de cálculo idónea, la relación coste/eficiencia de cada uno de los métodos y otros aspectos relevantes, como el tratamiento de los cambios de calidad de las viviendas.

Posteriormente, en la **segunda fase** del Grupo de Estudio (que se desarrolló en 2006 y 2007) se incorporaron al mismo otros siete países: Chipre, Eslovenia, Francia, Grecia, Italia, Países Bajos y República de Eslovaquia. El objetivo de esta fase era obtener datos sobre la evolución de los precios de la vivienda, tanto nueva como de segunda mano, en cada uno de estos países. Los resultados debían basarse en metodologías de cálculo comparables, que permitieran la construcción de un índice de precios de la vivienda en propiedad con el mayor grado de armonización. Esta es la principal razón por la que el proyecto de implantación del IPV debe tener como referencia las pautas generales marcadas por Eurostat para el indicador armonizado.

Además, cada país debía afrontar el cometido de calcular un índice de costes asociados a la adquisición de vivienda, estudiando y seleccionando para ello los artículos representativos de ese gasto.

Durante este periodo, en el INE se analizaron todas las fuentes disponibles para la obtención de la información necesaria para el cálculo del IPV, atendiendo por un lado

a la disponibilidad, puntualidad y contenido de la información; y por otro, a las exigencias de Eurostat al respecto, que requiere que los precios con los que se calculen los índices sean precios efectivamente pagados de las compraventas realizadas en el periodo de referencia del índice.

Cada una de las fuentes de información disponibles sobre el mercado de la vivienda ofrece una perspectiva distinta del proceso de compraventa. Además de las dos fuentes ya mencionadas (tasaciones y promociones inmobiliarias) existen otros agentes que proporcionan información sobre vivienda: escrituraciones, préstamos hipotecarios, Catastro y agencias inmobiliarias.

Después de un minucioso estudio de dichas fuentes y, tras sopesar la utilidad y conveniencia de cada una para los objetivos establecidos, se ha considerado que los valores de escrituración proporcionan la información más adecuada para el seguimiento de los precios de la vivienda en propiedad bajo los parámetros de comparabilidad y homogeneidad exigidos por Eurostat.

Los motivos por los que se han desestimado las demás fuentes disponibles para ser utilizadas en este proyecto son diversos. En el caso de los valores de tasación, el valor de la vivienda estimado por las compañías no tiene por qué ser el mismo que el precio de la transacción; asimismo, no todas las viviendas tasadas son objeto de compraventa, ni están consideradas las viviendas cuyo pago haya sido realizado en efectivo, sin crédito hipotecario.

Por su parte, los préstamos hipotecarios se han descartado porque los valores de los mismos pueden no coincidir con el precio de la transacción, y esta diferencia entre el precio efectivo y el valor del préstamo varía en cada caso. Desde el punto de vista práctico, dado que esta información procede de los registros notariales y de la propiedad, la disponibilidad en tiempo y forma es la misma.

También se ha prescindido de los datos del Catastro porque el valor catastral no proporciona una medida adecuada del precio de la vivienda, para los objetivos marcados por el IPV.

Por último, a pesar de que la recogida de precios a través de una encuesta dirigida a promotoras inmobiliarias podría ofrecer información útil y completa para la consecución del proyecto, ésta sólo cubriría la vivienda nueva, lo que obligaría a utilizar otras fuentes para la obtención de los precios de las viviendas de segunda mano. Además, la complejidad y coste de tal operación la hacen inviable.

De todas estas fuentes la finalmente 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 y se corresponden con el valor de la escrituración de la vivienda.

Además, la utilización de registros administrativos permite disponer de la información del total de la población que conforma el ámbito de estudio o investigación, lo que favorece la precisión de los resultados y reduce los costes si se compara con otras estadísticas que utilizan técnicas de muestreo para la recogida en campo de los datos.

A continuación se enumeran las características más destacables que presentan los valores de escrituración:

- **Disponibilidad.** A comienzos de 2005 el Consejo General del Notariado y el INE firmaron un Convenio de Colaboración para la realización de operaciones estadísticas relativas a bienes inmuebles, lo que permite disponer de la información necesaria para el cálculo del IPV (datos relativos a las transacciones de viviendas) con la periodicidad y plazos adecuados para la operación estadística en cuestión.
- **Puntualidad.** El INE recibe mensualmente la información de las transmisiones de bienes inmuebles que se producen en España con un desfase de aproximadamente seis semanas después de la finalización del trimestre de referencia de los datos. Este plazo se ha acordado con el objeto de incluir más del 90% del total de las transacciones efectuadas en el trimestre de referencia.
- **Actualidad.** Las transacciones entran en el cálculo del índice el mismo trimestre en que tienen lugar.
- **Cobertura.** Incluye información sobre las compraventas realizadas en el territorio nacional a lo largo del trimestre, tanto de viviendas nuevas como de segunda mano, e independientemente del modo de pago.
- **Contenido.** La base de datos de los notarios contiene información detallada que permite establecer tipologías de viviendas muy concretas. Las variables que se reciben se refieren a características de la vivienda, así como del comprador; entre las primeras destacan las relativas al tamaño del inmueble y a su localización.

Por otra parte, y de forma paralela al estudio de las fuentes, durante esta segunda fase se determinó el método de cálculo idóneo para el índice de precios de vivienda, que solucionara el problema de los cambios en la calidad y en la composición de la muestra de viviendas que cada trimestre se utilizan para calcular el índice. Este método requiere de la utilización de modelos de regresión, tal y como se verá detalladamente en el apartado dedicado a la metodología de cálculo.

Por último, en diciembre de 2007 comenzó la tercera fase del Grupo de Estudio, en la que participan 26 países de la UE y entre cuyos cometidos se encuentra la producción periódica de índices y la mejora en su metodología de cálculo. Además, se continuará calculando el índice de los costes asociados a la adquisición de vivienda, y se comenzará a calcular un índice de grandes reparaciones.

3. Objetivos

El IPV tiene como principal objetivo medir la **evolución del nivel de los precios de compraventa de las viviendas de precio libre**, nuevas y de segunda mano, a lo largo del tiempo. Se trata, por tanto, de un indicador concebido únicamente para establecer comparaciones en el tiempo.

No entra dentro del ámbito del mismo la medición de los niveles de precios. Por tanto, no se podrán establecer comparaciones espaciales de los niveles de precios, aunque sí de sus evoluciones.

De acuerdo con los criterios sobre cobertura del Índice de Precios de Consumo Armonizado (IPCA), las viviendas protegidas se excluyen del cálculo del IPV porque no son accesibles a todos los posibles compradores y no se rigen por los mecanismos habituales del mercado. Se trata de viviendas cuya tipología, dimensiones y precios están reglamentadas por la Administración, como condición para poder acogerse a determinadas ventajas tanto económicas como fiscales por parte de los compradores, los cuales a su vez deben reunir unas condiciones establecidas en cuanto a titularidad de inmuebles, ingresos familiares, etc.

Asimismo, dentro del ámbito de la producción de estadísticas armonizadas de la Unión Europea, el IPV tiene como objetivo servir de elemento de comparación entre los Estados Miembros de la UE, en lo que se refiere a la evolución de los precios de la vivienda. En este sentido, se ha concebido bajo los mismos conceptos y metodología que se utiliza en la producción de los IPCA de la UE.

4. Ámbitos de investigación

Unidades de análisis

Teniendo en cuenta que el objetivo del IPV es medir la evolución del nivel de los precios de compraventa de las viviendas de precio libre en España, la unidad de análisis es la vivienda libre.

Ámbito poblacional

La población o estrato de referencia del IPV incluye a toda la población (personas físicas), tanto residente como no residente en España, que haya adquirido una vivienda en el periodo de referencia. Las compras efectuadas por personas jurídicas o entidades financieras no forman parte del ámbito poblacional del IPV.

Ámbito geográfico

El ámbito geográfico de la investigación lo constituye todo el territorio nacional.

Ámbito temporal

El IPV tiene una periodicidad trimestral, lo que permite estimar la evolución de precios entre trimestres consecutivos, la acumulada a lo largo del año, así como la anual.

5. Fuente de información. Variables utilizadas

Como se ha comentado anteriormente, la información utilizada para el cálculo del IPV proviene del Consejo General del Notariado, con quien el INE firmó un convenio de colaboración destinado a facilitar la explotación de los datos de los notarios con fines estadísticos. Conforme a ello, el Consejo General del Notariado, a través de la Agencia Notarial de Certificación (ANCERT), suministra los datos que constituyen la fuente principal de información de este indicador.

ANCERT proporciona mensualmente la información de las transmisiones de bienes inmuebles que se producen en España, además de sucesivas actualizaciones en las que se incorporan nuevas observaciones o modificaciones de las enviadas previamente.

El IPV es un índice coyuntural, de periodicidad trimestral, y en su elaboración intervienen los últimos datos con los que cuenta el INE en el momento del cálculo, los cuales representan, en media, más del 90% del total de transmisiones efectuadas cada trimestre. Esta cobertura satisface sobradamente las necesidades de este indicador, teniendo en cuenta los criterios de puntualidad y oportunidad exigibles a una estadística de estas características. No obstante, se ha comprobado que la información que se recibe con posterioridad al cálculo del indicador tiene un efecto despreciable en el resultado final.

La información remitida mensualmente por ANCERT incluye las siguientes variables:

- **VARIABLES DE LOCALIZACIÓN DE LA VIVIENDA:**
 - o Comunidad Autónoma.
 - o Provincia.
 - o Municipio.
 - o Código postal.
 - o Tipo, nombre y número de la vía.
 - o Duplicado, bloque, escalera, planta y puerta.
- **VARIABLE DE CARÁCTER TEMPORAL QUE INDICA EL DÍA EN QUE SE REALIZA LA TRANSMISIÓN DE LA VIVIENDA:**
 - o Fecha de autorización.
- **CÓDIGOS DE LOS ACTOS JURÍDICOS.** Se reciben las transacciones correspondientes a los siguientes actos jurídicos:
 - o Compraventa de inmuebles.
 - o Adjudicación de cooperativa de vivienda a sus socios.
 - o Adjudicación a comunero en la comunidad de promoción inmobiliaria.

– **VARIABLES RELATIVAS AL PRECIO DE LA VIVIENDA:**

- Precio de la operación.
- Valor del objeto.

En los actos con pluralidad de objetos, es decir, cuando en un mismo acto jurídico se transmiten varios objetos (por ejemplo, una vivienda y un aparcamiento) y en el *precio de la operación* están incluidos todos ellos, se consigna el valor de cada uno de ellos en el campo *valor del objeto*.

– **VARIABLES RELATIVAS A LAS CARACTERÍSTICAS DE LA VIVIENDA:**

- Clase de inmueble. Indica el tipo de finca urbana; se distingue entre viviendas, aparcamientos, trasteros y terrenos o solares.
- Referencia catastral del inmueble, o motivo de la ausencia o razón por la cual no se ha podido obtener, en su caso.
- Tipo de vivienda, distinguiendo entre piso y vivienda unifamiliar.
- Variable que indica si la vivienda es nueva o de segunda mano. Se considera que una vivienda es nueva cuando es la primera transmisión en la escritura de compraventa, normalmente realizada por el promotor o constructor a favor del primer adquiriente; en el resto de transmisiones, es decir, cuando existe más de una transmisión en la escritura pública, la vivienda se considera de segunda mano.
- Superficie construida en m².
- Variables que indican si el aparcamiento y el trastero están incluidos en el precio.
- Variable que indica si la vivienda es de precio libre o de protección oficial.

– **VARIABLES RELATIVAS AL COMPRADOR:**

- Tipo de persona. Se reciben las transacciones efectuadas por personas físicas, quedan excluidas las realizadas por personas jurídicas o entidades financieras.
- País, provincia y municipio de residencia del adquiriente.

Las variables relativas al comprador permiten la adaptación a la cobertura poblacional del IPV, idéntica a la del IPCA, que debe incluir las compras realizadas por particulares en el territorio nacional, independientemente de su nacionalidad.

Además de la información recibida del Consejo General del Notariado, se han generado otras variables, a partir de información procedente de otras fuentes, que permiten aumentar la capacidad explicativa del modelo de regresión utilizado en la estimación de precios. Estas variables son las siguientes:

- **Cluster de provincias.** Agrupación de las 52 provincias en 6 grupos mediante la aplicación de un análisis cluster, utilizando como variable de clasificación el importe anual medio de la hipoteca, obtenido de la Estadística de Hipotecas del INE considerando únicamente las fincas urbanas, y más concretamente las viviendas.
- **Tamaño del municipio.** Clasificación de los municipios distinguiendo entre capitales de provincia y municipios no capitales grandes (con más de 100.000 habitantes), medianos (de 10.000 a 100.000 habitantes) y pequeños (menos de 10.000 hab.), utilizando los últimos datos de población disponibles del Padrón Continuo de Habitantes (INE).
- **Municipio turístico.** Se obtiene a partir de la información sobre el número de pernoctaciones anuales realizadas en cada tipo de establecimiento turístico, facilitada por las encuestas de Ocupación Hotelera, de Ocupación en Apartamentos Turísticos y de Ocupación en alojamientos de Turismo Rural (INE).
- **Tipo de entorno.** Clasificación de los códigos postales en 14 categorías, a partir de la información del Censo de Población y Viviendas de 2001 sobre la condición socioeconómica media y el nivel medio de problemas, considerando como problemas aspectos tales como la delincuencia, el nivel de ruido o de suciedad en la zona, las malas comunicaciones, la escasez de zonas verdes, etc. También se ha tenido en cuenta el propio precio del metro cuadrado por código postal, en un periodo anual anterior.

6. Metodología de cálculo

El sistema de cálculo del IPV está basado en la combinación de dos elementos básicos que reflejan las características del mercado inmobiliario, y que son esenciales en el cálculo de los índices de precios: los precios de las viviendas, que representan la confluencia de la oferta y la demanda del mercado, y las ponderaciones, o importancia relativa de cada tipología de vivienda según el valor de compra.

La combinación de estos dos elementos para obtener el IPV se realiza mediante la **fórmula del índice de Laspeyres encadenado**, la misma que se emplea en el cálculo del IPC/IPCA.

Además de la consideración de los dos elementos citados anteriormente, otro de los aspectos 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. Cuando los precios observados corresponden a viviendas, este aspecto es de máxima importancia. En este caso, el seguimiento del precio de una misma vivienda a lo largo del tiempo resulta imposible; de hecho, la composición de la cesta de viviendas utilizada para el cálculo del índice es diferente cada trimestre, ya que está compuesta por las viviendas vendidas en dicho trimestre. Por tanto, si no se ajustan los precios por el cambio en la composición de la muestra o la calidad de las viviendas, la medición de la evolución de los mismos no sería correcta.

Una posible solución es la creación de estratos de viviendas que agrupen aquellas con características similares, y seguir el precio medio de cada estrato en lugar del precio de cada vivienda individualmente para, posteriormente, obtener la evolución conjunta ponderando cada uno de los estratos. En este sentido, es conveniente crear estratos pequeños; ya que cuanto más definida esté la tipología de vivienda, mejor será el ajuste por cambio de calidad y composición.

Por otro lado, para obtener precios medios representativos de cada estrato con los estimadores tradicionales, es necesario disponer de un número mínimo de observaciones por estrato cada trimestre. Esta exigencia obligaría a disminuir el grado de detalle del estrato, reduciendo el número de características que lo definen. Como consecuencia, las viviendas pertenecientes a un mismo estrato podrían no ser tan homogéneas como sería deseable. Por ello, el IPV utiliza modelos de regresión hedónica que permiten la estimación trimestral de precios por estrato, independientemente de que existan o no viviendas pertenecientes al mismo en el trimestre. De esta forma, el número de tipologías consideradas y el grado de detalle en la definición de éstas es mayor, lo que hace mejorar significativamente el ajuste.

A continuación se presenta con mayor detalle el proceso de cálculo del IPV, incidiendo en los aspectos señalados y, de forma especial, en el modelo de regresión utilizado.

6.1 Precios

Como ya se ha dicho anteriormente, los precios por metro cuadrado son uno de los elementos básicos en el cálculo de este indicador. Sin embargo, dada la heterogeneidad de las viviendas, es preciso aplicar a estos precios un proceso que garantice la comparabilidad de los mismos; por tanto, los precios que finalmente intervienen en el cómputo del IPV son los que se obtienen para cada estrato o tipología de vivienda después de la aplicación del proceso de estimación. Este proceso utiliza la información original sobre compra-ventas de viviendas de cada trimestre a la que se le aplica un modelo de regresión.

El proceso de validación y depuración de los datos para el cálculo del IPV, se realiza en dos etapas. Por un lado, antes de la aplicación del modelo de regresión, se analiza la información y se detectan y corrigen los valores atípicos encontrados. Por otro lado, el propio modelo de regresión detecta las observaciones atípicas, que posteriormente se corrigen, basándose en la información conjunta que proporcionan todas las variables explicativas del modelo.

Así pues, el modelo de regresión inicial se aplica sobre el fichero de datos que resulta de la primera fase de validación y depuración, y de este modo se detectan y eliminan los atípicos del modelo.

Para los años 2007 y 2008 se han construido en torno a 52.000 tipologías de vivienda diferentes, que se han determinado observando las características físicas y de localización de las viviendas objeto de compraventa durante tres años: 2005, 2006 y 2007. Los precios de cada tipología se estiman trimestralmente a partir de las compraventas de viviendas realizadas en el trimestre, independientemente del número de transacciones trimestrales de cada tipología.

El modelo permite, por tanto, obtener un precio estimado para todas las tipologías de vivienda, independientemente del número de transacciones realizadas en ese trimestre en ese estrato, a partir de la información suministrada por la muestra trimestral de compraventas. Finalmente, el precio medio trimestral de cada tipología se obtiene multiplicando el vector de parámetros (que varía trimestralmente) con el vector de características que define cada tipología de vivienda. Dichos parámetros recogen los precios implícitos de estas características.

Con los precios estimados por el modelo se calculan los índices elementales de cada tipología, que junto a su ponderación se utilizan para el cálculo de los índices agregados.

6.2 Modelo de regresión

En el modelo de regresión que se aplica para el cálculo del IPV las variables explicativas, que recogen las características físicas y de localización de la vivienda, son todas categóricas, es decir, toman un número finito de valores. Cada combinación posible de los valores de estas variables conforma lo que se ha denominado celda.

Este modelo tiene como variable dependiente el logaritmo neperiano del precio por metro cuadrado de la vivienda; de este modo se capta la relación no lineal entre el precio y el resto de variables, y además permite una interpretación sencilla de los coeficientes o parámetros del modelo.

El modelo que se ha utilizado para la estimación de precios en los años 2007 y 2008, incluye los siguientes efectos principales o variables explicativas:

- Nueva/Segunda mano.
- Tipo de vivienda (piso/unifamiliar).
- Aparcamiento (Sí/No).
- Trastero (Sí/No).
- Cooperativa (Sí/No).
- Superficie (en intervalos).
- Cluster de provincias.
- Tamaño del municipio.
- Municipio turístico (Sí/No).
- Tipo de entorno.

Las cinco primeras son variables dicotómicas que se obtienen directamente del fichero de datos de los notarios; las cuatro últimas se han creado a partir de la información de otras fuentes, tal como se ha descrito en el apartado 5. Por último, la variable superficie se ha formado clasificando en 10 intervalos los valores de la variable numérica *superficie construida de la vivienda*:

1. $< 40 \text{ m}^2$
2. $[40, 60) \text{ m}^2$
3. $[60, 75) \text{ m}^2$
4. $[75, 90) \text{ m}^2$
5. $[90, 105) \text{ m}^2$
6. $[105, 120) \text{ m}^2$
7. $[120, 150) \text{ m}^2$
8. $[150, 180) \text{ m}^2$
9. $[180, 240) \text{ m}^2$
10. $\geq 240 \text{ m}^2$

Además, el modelo incluye las interacciones dobles más significativas entre estos efectos principales. Los criterios seguidos para la selección de las interacciones son tres: deben ser significativas, su aportación al poder explicativo del modelo debe ser lo más alta posible, y el número de observaciones trimestrales debe ser superior a 50 en cada una de las combinaciones o pares de valores posibles de la interacción.

El modelo utilizado para los datos del año 2008 consta de 9 interacciones, que junto a los efectos principales suman un total de 157 parámetros a estimar cada trimestre.

Tanto los efectos principales como las interacciones podrán cambiar anualmente, ya que el modelo será objeto de revisión cada año. Esta revisión consistirá en:

- actualizar con la última información disponible de las fuentes utilizadas en su elaboración, las variables cluster de provincias, tamaño del municipio, municipio turístico y tipo de entorno, lo que puede hacer variar el número de categorías de las mismas;
- paralelamente, se podrán incorporar al modelo nuevas variables explicativas provenientes de la base de datos de los notarios o bien creadas a partir de información complementaria.

En el anexo II se incluye la especificación del modelo de regresión, la corrección de heterocedasticidad y la estimación de los precios por celda.

6.3 Ponderaciones

La estructura de ponderaciones permite establecer la importancia o peso que cada estrato o tipología de vivienda tiene frente a todas las demás, en función del gasto efectuado en la compra de cada tipo de vivienda respecto al gasto total en compra de viviendas durante el periodo de referencia.

Se trata, pues, de una variable de flujo (transacciones realizadas) cuya composición puede cambiar de año en año, y no de una variable de stock como sería, por ejemplo, el número de viviendas en propiedad existentes en España.

La fuente de información utilizada para la obtención de las ponderaciones es la misma de la que se obtienen los precios, ya que los datos notariales permiten conocer los tipos de transacciones realizadas a lo largo del tiempo, tanto en número como en valor.

La estructura de ponderaciones se obtiene con la información referida a los tres últimos años disponibles. Con ello se persiguen dos objetivos:

- Garantizar la estabilidad en la estructura de ponderaciones, ya que si ésta cambiara significativamente de un año a otro, se podrían ver afectadas las tasas anuales.
- Mejorar la representatividad del índice. Cuantos más años intervienen en el cálculo, mayor es la variedad de tipologías o estratos de viviendas representadas y, por tanto, mejor es el ajuste por cambio de calidad y composición.

Debido al dinamismo del mercado inmobiliario, es conveniente actualizar anualmente la estructura de ponderaciones de forma que ésta represente la situación del mercado lo más fielmente posible. Esto se va a poder llevar a cabo ya que el IPV es un índice de Laspeyres encadenado.

Como ya se ha comentado anteriormente, la ponderación de cada estrato representa la relación entre el gasto que los hogares realizan en la compra de viviendas de la misma tipología, es decir, pertenecientes al mismo estrato, y el gasto total realizado en compra de vivienda. Para el cálculo de las ponderaciones se utiliza el gasto estimado, calculado multiplicando el precio del m² de las viviendas incluidas en el estrato e del 4º trimestre del año anterior al corriente ($\hat{P}_e^{4, a-1}$), por los m² vendidos de viviendas pertenecientes a dicho estrato en el periodo de referencia:

$$W_e^a = \frac{Q_e^{(a-1, a-2, a-3)} \times \hat{P}_e^{4, a-1}}{\sum_{\forall e} Q_e^{(a-1, a-2, a-3)} \times \hat{P}_e^{4, a-1}} \quad a \geq 2008$$

siendo,

$Q_e^{(a-1, a-2, a-3)}$ la cantidad anual media de metros cuadrados de viviendas pertenecientes al estrato e , vendidas en el periodo de referencia de las ponderaciones ($a-1, a-2, a-3$),

W_e^a la ponderación del estrato e , en tanto por uno, vigente en el año a .

El motivo por el cual el gasto en vivienda se calcula utilizando los precios estimados por el modelo de regresión para el último trimestre del año anterior, es corregir el desfase entre el periodo de referencia de las cantidades (los años $a-3, a-2, a-1$) y el periodo de referencia de los precios (el 4º trimestre de $a-1$). De este modo, utilizando la última información disponible sobre precios, se actualiza el gasto de cada tipología de vivienda al último trimestre del año anterior, que es también el de referencia de los precios.

El hecho de utilizar precios estimados en lugar de precios recogidos es debido a que en el periodo de referencia (el último trimestre del año $a-1$) posiblemente no se hayan vendido viviendas de todas las tipologías y, por tanto, no se disponga de información de precios transados para todos los estratos.

La ponderación de cualquier agregado A , ya sea funcional o geográfico, se obtiene como suma de las ponderaciones de los estratos que comprende dicho agregado:

$$W_A^a = \sum_{e \in A} W_e^a$$

6.4 Cálculo de índices

La fórmula general utilizada para el cálculo del IPV es un índice de Laspeyres encadenado, análoga a la utilizada en el Índice de Precios de Consumo (IPC). En el caso del IPV, por tratarse de un indicador trimestral, el periodo utilizado para el encadenamiento es el cuarto trimestre de cada año y no el mes de diciembre.

La utilización de índices encadenados permite la actualización anual de las ponderaciones, así como la posibilidad de realizar cambios metodológicos (revisión del modelo, inclusión de nuevos estratos, etc.), a diferencia de lo que sucede con un índice de Laspeyres de base fija, en el que tanto las ponderaciones como la metodología se mantienen fijas durante todo el periodo de vigencia de la base.

En un índice encadenado se definen tres periodos de referencia:

- **Periodo de referencia del índice o periodo base.** Es aquel en el que todos los índices se hacen igual a 100. Normalmente se trata de un periodo anual. En el IPV el año base es 2007, y todos los índices que se publiquen a partir del mismo tendrán como referencia dicho periodo.
- **Periodo de referencia de las ponderaciones.** Es el periodo al que se refieren los datos utilizados en el cálculo de ponderaciones.

Cada año se actualizarán las ponderaciones del IPV con la última información disponible sobre las compraventas de viviendas efectuadas en los tres últimos años disponibles, actualizada a precios del 4º trimestre del año anterior al corriente. Consecuentemente, el periodo de referencia de las ponderaciones variará anualmente.

En el año 2008, el periodo de referencia de las ponderaciones será el constituido por los años 2005, 2006 y 2007, actualizado a precios del 4º trimestre del 2007.

- **Periodo de referencia de los precios.** Es el periodo con cuyos precios se comparan los precios corrientes; es decir, el periodo elegido para el cálculo de los índices simples. Se trata del cuarto trimestre del año inmediatamente anterior al corriente, por lo que variará cada año.

A continuación se detalla la fórmula general de cálculo de los índices elementales y de los índices agregados, así como el sistema general de cálculo de los índices encadenados.

ÍNDICES ELEMENTALES

Un agregado elemental es el componente de más bajo nivel de agregación para el cual se obtienen índices, y en cuyo cálculo no intervienen ponderaciones; a los índices de estos agregados se les llama índices elementales. En el IPV el agregado elemental es el estrato que recoge una misma tipología de vivienda.

El índice elemental del estrato se obtiene como cociente del precio estimado por el modelo para las viviendas pertenecientes a dicho estrato en el periodo actual, y el precio estimado del cuarto trimestre del año anterior:

$${}_{(4, a-1)}I_e^{q, a} = \frac{\hat{P}_e^{q, a}}{\hat{P}_e^{4, a-1}} \times 100 \quad a \geq 2008$$

siendo,

$\hat{P}_e^{q, a}$ el precio estimado del metro cuadrado para las viviendas pertenecientes al estrato e , en el trimestre q del año a , y

$\hat{P}_e^{4, a-1}$ el precio estimado del metro cuadrado para las viviendas del estrato e , en el 4º trimestre del año $(a-1)$. La estimación de este precio ha sido realizada con el mismo modelo de regresión¹ empleado en la estimación del precio del numerador.

ÍNDICES AGREGADOS REFERIDOS AL CUARTO TRIMESTRE

El cálculo del índice de una agregación A , ya sea de tipo funcional o geográfica, se realiza a partir de los índices elementales de los estratos pertenecientes a dicha agregación y sus correspondientes ponderaciones, de acuerdo con la siguiente expresión:

$${}_{(4, a-1)}I_A^{q, a} = \sum_{e \in A} W_e^a \times {}_{(4, a-1)}I_e^{q, a} \quad a \geq 2008$$

siendo,

W_e^a la ponderación del estrato e , en tanto por uno, vigente durante el año a , y

${}_{(4, a-1)}I_e^{q, a}$ el índice elemental del estrato e , en el trimestre q del año a .

A partir de los índices referidos al cuarto trimestre se calculan las repercusiones trimestrales y acumuladas a lo largo del año.

ÍNDICES EN BASE 2007

Los índices en base 2007 son los que se publican y se obtienen encadenando los índices referidos al cuarto trimestre del año anterior, según la siguiente expresión:

$$\begin{aligned} {}_{07}I_A^{q, a} &= {}_{07}I_A^{4, (a-1)} \times \left(\frac{{}_{4, (a-1)}I_A^{q, a}}{100} \right) = \\ &= {}_{07}I_A^{4, 07} \times \left(\frac{{}_{(4, 07)}I_A^{4, 08}}{100} \right) \times \dots \times \left(\frac{{}_{(4, a-2)}I_A^{(4, a-1)}}{100} \right) \times \left(\frac{{}_{(4, a-1)}I_A^{q, a}}{100} \right) \end{aligned} \quad a \geq 2008$$

¹ El modelo de regresión se revisa cada año, de modo que los precios para el cuarto trimestre de cada año a se han de estimar de dos formas diferentes. Por un lado; se utilizará el modelo vigente en el año a para calcular el numerador de los índices elementales del cuarto trimestre del año a . Por otro, con el modelo revisado, vigente en el año posterior $a+1$, se calcularán los denominadores de los índices elementales de los cuatro trimestres del año $a+1$.

Para el año base, 2007, los índices se definen como sigue a continuación:

$${}_{07}I_A^{q,07} = \frac{\sum_{e \in A} Q_e^{(05,06,07)} \times \hat{P}_e^{q,07}}{\sum_{e \in A} Q_e^{(05,06,07)} \times \bar{\hat{P}}_e^{07}} \times 100$$

con $\bar{\hat{P}}_e^{07} = \frac{1}{4} \sum_{q=1}^4 \hat{P}_e^{q,07}$; por tanto: $\frac{1}{4} \sum_{q=1}^4 I_e^{q,07} = 100$

A partir de los índices en base 2007 se obtienen las tasas de variación trimestral, la acumulada (o en lo que va de año) y la anual. Las tasas de variación trimestral y acumulada también se pueden obtener a partir de los índices referidos al 4º trimestre, como se verá a continuación.

6.5 Cálculo de tasas de variación

TASA DE VARIACIÓN TRIMESTRAL

La tasa de variación trimestral de un índice en el periodo (q, a) se calcula como el cociente entre el índice del trimestre corriente q y el índice del trimestre anterior q-1, ambos en base 2007, según la siguiente fórmula:

$$\Delta^{qa/(q-1)a} = \left(\frac{{}_{07}I^{qa}}{{}_{07}I^{(q-1)a}} - 1 \right) \times 100$$

siendo:

$\Delta^{qa/(q-1)a}$ la tasa de variación trimestral de precios en el trimestre q del año a,

${}_{07}I^{qa}$ el índice del trimestre q del año a, en base 2007, y

${}_{07}I^{(q-1)a}$ el índice del trimestre q-1 del año a, en base 2007.

Estas tasas también se pueden calcular con los índices referidos al cuarto trimestre del año inmediatamente anterior, obteniendo el mismo resultado que con los índices en base 2007:

$$\Delta^{qa/(q-1)a} = \left(\frac{{}_{07}I^{qa}}{{}_{07}I^{(q-1)a}} - 1 \right) \times 100 = \left(\frac{{}_{07}I^{4(a-1)} \times \frac{{}_{(4,a-1)}I^{qa}}{100}}{{}_{07}I^{4(a-1)} \times \frac{{}_{(4,a-1)}I^{(q-1)a}}{100}} - 1 \right) \times 100 =$$

$$= \left(\frac{{}_{(4,a-1)}I^{qa}}{{}_{(4,a-1)}I^{(q-1)a}} - 1 \right) \times 100$$

TASA DE VARIACIÓN ACUMULADA

La tasa de variación acumulada, o en lo que va de año, se calcula como cociente entre el índice del trimestre corriente y el índice del cuarto trimestre del año anterior, ambos en base 2007:

$$\Delta^{qa/4(a-1)} = \left(\frac{{}_{07}I^{qa}}{{}_{07}I^{4(a-1)}} - 1 \right) \times 100$$

siendo:

$\Delta^{qa/4(a-1)}$ la tasa de variación acumulada de precios en el trimestre q del año a ,

${}_{07}I^{qa}$ el índice en base 2007, en el trimestre q del año a , y

${}_{07}I^{4(a-1)}$ el índice en base 2007, en el cuarto trimestre del año $a-1$.

Al igual que ocurre con las tasas trimestrales, las acumuladas también se pueden calcular con los índices referidos al cuarto trimestre del año anterior, obteniendo el mismo resultado que con los índices en base 2007.

TASA DE VARIACIÓN ANUAL

La tasa de variación anual se calcula como el cociente entre los índices publicados del trimestre corriente y del mismo trimestre del año anterior, ambos en base 2007:

$$\Delta^{qa/q(a-1)} = \left(\frac{{}_{07}I^{qa}}{{}_{07}I^{q(a-1)}} - 1 \right) \times 100$$

siendo:

$\Delta^{qa/q(a-1)}$ la tasa de variación anual de precios en el trimestre q del año a ,

${}_{07}I^{qa}$ el índice en base 2007, en el trimestre q del año a , y

${}_{07}I^{q(a-1)}$ el índice en base 2007, en el trimestre q del año $a-1$.

6.6 Cálculo de repercusiones

REPERCUSIONES TRIMESTRALES

La repercusión de la variación trimestral de un estrato o conjunto de estratos de viviendas en el índice general, se define como la parte de la variación trimestral del índice general que corresponde a dicho estrato o conjunto de estratos. Por tanto, la suma de las repercusiones trimestrales de todos los estratos de viviendas incluidos en el IPV, es igual a la variación trimestral del índice general.

En otras palabras, la repercusión que la variación trimestral de precios de un estrato o conjunto de estratos tiene en la variación trimestral del índice general, es la variación que éste hubiera experimentado si todos los precios del resto de estratos no hubieran variado en ese trimestre.

La fórmula de la repercusión trimestral de un estrato (o de un conjunto de estratos) determinado, en el trimestre q del año a , es la siguiente:

$$R_e^{qa/(q-1)a} = \frac{{}_{4(a-1)}I_e^{qa} - {}_{4(a-1)}I_e^{(q-1)a}}{{}_{4(a-1)}I_G^{(q-1)a}} \times W_e^a \times 100$$

donde:

${}_{4(a-1)}I_e^{qa}$ es el índice referido al 4º trimestre del año $(a-1)$ del estrato e , en el trimestre q del año a ,

${}_{4(a-1)}I_e^{(q-1)a}$ es el índice referido al cuarto trimestre del año $(a-1)$ del estrato e , en el trimestre $(q-1)$ del año a ,

${}_{4(a-1)}I_G^{(q-1)a}$ es el índice general referido al cuarto trimestre del año $(a-1)$, en el trimestre $q-1$ del año a , y

W_e^a es la ponderación vigente en el año a del estrato e , en tanto por uno.

Como se puede observar, las repercusiones se calculan a partir de los índices referidos al cuarto trimestre del año anterior (índices no publicados). Una expresión alternativa de la fórmula anterior es la que sigue:

$$\begin{aligned} R_e^{qa/(q-1)a} &= \frac{{}_{4(a-1)}I_e^{qa} - {}_{4(a-1)}I_e^{(q-1)a}}{{}_{4(a-1)}I_G^{(q-1)a}} \times W_e^a \times 100 = \\ &= \frac{{}_{4(a-1)}I_e^{qa} - {}_{4(a-1)}I_e^{(q-1)a}}{{}_{4(a-1)}I_G^{(q-1)a}} \times \frac{{}_{4(a-1)}I_e^{(q-1)a}}{{}_{4(a-1)}I_e^{(q-1)a}} \times W_e^a \times 100 = \\ &= \Delta_e^{qa/(q-1)a} \times \frac{{}_{4(a-1)}I_e^{(q-1)a}}{{}_{4(a-1)}I_G^{(q-1)a}} \times W_e^a \end{aligned}$$

Por tanto, la repercusión trimestral de un estrato determinado e , es el producto de su tasa de variación trimestral de precios, $\Delta_e^{qa/(q-1)a}$, por su ponderación, W_e^a , y por el cociente entre el índice del estrato y el índice general del trimestre anterior,

$$\frac{{}_{4(a-1)}I_e^{(q-1)a}}{{}_{4(a-1)}I_G^{(q-1)a}} .$$

Como ya se ha mencionado anteriormente, la suma de las repercusiones trimestrales de todos los estratos que componen la cesta de tipologías de viviendas del IPV es la variación trimestral del índice general.

$$\begin{aligned}
\sum_e R_e^{qa/(q-1)a} &= \sum_e \frac{4(a-1) I_e^{qa} - 4(a-1) I_e^{(q-1)a}}{4(a-1) I_G^{(q-1)a}} \times W_e^a \times 100 = \\
&= \left(\sum_e \frac{4(a-1) I_e^{qa}}{4(a-1) I_G^{(q-1)a}} \times W_e^a - \sum_e \frac{4(a-1) I_e^{(q-1)a}}{4(a-1) I_G^{(q-1)a}} \times W_e^a \right) \times 100 = \\
&= \left(\frac{\sum_e 4(a-1) I_e^{qa} \times W_e^a - \sum_e 4(a-1) I_e^{(q-1)a} \times W_e^a}{4(a-1) I_G^{(q-1)a}} \right) \times 100 = \\
&= \frac{4(a-1) I_G^{qa} - 4(a-1) I_G^{(q-1)a}}{4(a-1) I_G^{(q-1)a}} \times 100 = \Delta_G^{qa/(q-1)a}
\end{aligned}$$

REPERCUSIONES ACUMULADAS

La repercusión acumulada o en lo que va de año de un estrato o conjunto de estratos en el índice general, representa la variación acumulada que experimentaría el índice general si el resto de estratos no hubieran sufrido variación alguna de precios en lo que va de año; o lo que es lo mismo, es la parte de la variación acumulada debida a dicho estrato o conjunto de estratos.

La fórmula de la repercusión acumulada o en lo que va de año de un estrato determinado e (o de una agregación determinada) en el trimestre q del año a, es la siguiente:

$$\begin{aligned}
R_e^{qa/4(a-1)} &= \frac{4(a-1) I_e^{qa} - 4(a-1) I_e^{4(a-1)}}{4(a-1) I_G^{4(a-1)}} \times W_e^a \times 100 = \\
&= \frac{4(a-1) I_e^{qa} - 100}{100} \times W_e^a \times 100 = \left(4(a-1) I_e^{qa} - 100 \right) \times W_e^a = \Delta_e^{qa/4(a-1)} \times W_e^a
\end{aligned}$$

donde:

$\Delta_e^{qa/4(a-1)}$ es la tasa de variación acumulada del estrato e , en el trimestre q del año a , y

W_e^a es la ponderación, del estrato e , vigente en el año a , en tanto por uno.

Por tanto, la repercusión en lo que va de año es el producto de la tasa de variación acumulada (en tanto por cien) y la ponderación (en tanto por uno).

En este caso, la suma de las repercusiones acumuladas de todos los estratos es igual a la variación en lo que va de año del índice general:

$$\begin{aligned} \sum_i R_e^{qa/4(a-1)} &= \sum_e \left({}_{4(a-1)}I_e^{qa} - 100 \right) \times W_e^a = \\ &= \sum_e {}_{4(a-1)}I_e^{qa} \times W_e^a - 100 \sum_e W_e^a = {}_{4(a-1)}I_G^{qa} - 100 = \\ &= \frac{{}_{4(a-1)}I_G^{qa} - 100}{100} \times 100 = \Delta_G^{qa/4(a-1)} \end{aligned}$$

7. Difusión

Se publican datos para el conjunto nacional, las 17 comunidades autónomas y las Ciudades Autónomas de Ceuta y de Melilla, lo que permite establecer comparaciones entre la evolución de los precios en las distintas regiones.

En cuanto a la desagregación funcional, se facilita información desagregada sobre vivienda nueva y vivienda de segunda mano, a nivel nacional.

En los próximos años, se valorará la posibilidad de ofrecer otro tipo de información más desagregada, según el tipo de vivienda o las características de las mismas.

Trimestralmente, se publican los siguientes datos:

- índices en base 2007;
- tasas de variación trimestral;
- tasas de variación en lo que va de año;
- tasas de variación anual;
- repercusiones trimestrales (de vivienda nueva y de segunda mano);
- repercusiones en lo que va de año (de vivienda nueva y de segunda mano).

Anexo I. Glosario de términos

- **ANCERT.** Agencia Notarial de Certificación, antes Instituto Notarial para las Tecnologías de la Información (INTI). Es una empresa del Consejo General del Notariado, constituida para la modernización tecnológica del notariado español.
- **Celda.** Combinación de los posibles valores de las variables o características (efectos principales del modelo de regresión) que definen un tipo concreto de vivienda.
- **Consejo General del Notariado.** Entidad que coordina los Colegios Notariales del territorio español. Gestiona la base de datos relativa a operaciones de bienes inmuebles (índice informatizado de los notarios), que se utiliza para el cálculo del IPV.
- **Cooperativa de vivienda.** Es el conjunto de personas que, cumpliendo con los requisitos generales de la cooperativa (redacción de unos estatutos, inscripción en el Registro de Sociedades Cooperativas, formación de los órganos por los que se rige, contabilidad, etc.), se unen para participar en un proyecto común, llevando a cabo cuantas actividades sean necesarias (búsqueda de solares, búsqueda de entidad financiera que financie la construcción, encargo al arquitecto, redacción de contratos de incorporación, contrato de obra, contratos de adjudicación de vivienda, etc.) para conseguir alojamientos y/o locales e instalaciones complementarias, para sí o para las personas que con ellos convivan.
- **Efecto principal.** Variable explicativa del modelo de regresión.
- **Hipoteca.** Derecho que contrae el prestamista frente al prestatario en caso de impago de las obligaciones de este último, y que se ejerce sobre el bien que aparece como garantía o colateral. En el caso de un préstamo hipotecario para una vivienda, la finca hipotecada suele ser la vivienda comprada.
- **Interacción.** Variable explicativa del modelo de regresión, obtenida como combinación de otras variables explicativas (efectos principales) del modelo.
- **Modelo de regresión hedónica.** Los modelos de precios hedónicos analizan el precio de un bien en función de sus múltiples características, por medio de la estimación del precio implícito de cada una de ellas.
- **Pisos.** Son las viviendas que forman parte de un edificio de dos o más plantas o alturas y tienen un acceso común a todas ellas desde la vía pública. Siempre que haya zonas privativas y zonas comunes existe una especial forma de copropiedad establecida como propiedad horizontal.
- **Tasación.** Una tasación es una estimación del valor de mercado de un bien basada en distintos parámetros que lo determinan; en el caso de las viviendas estos parámetros pueden ser la superficie, la ubicación, la antigüedad, etc. La mayoría de las tasaciones de vivienda se realizan a petición de una entidad bancaria con la finalidad de la concesión de un préstamo hipotecario destinado a la compra de la vivienda, y las suelen realizar empresas de tasación.

– **Valor de escrituración.** Escriturar es hacer constar con escritura pública y en forma legal un otorgamiento o un hecho.

El valor de escrituración de una vivienda es el que consta como valor de la vivienda en la escritura pública de compraventa y es, por tanto, el precio oficial de la misma.

– **Vivienda.** Todo recinto estructuralmente separado e independiente que, por la forma en que fue construido, reconstruido, transformado o adaptado, está concebido para ser habitado por personas y forma parte de un edificio.

– **Vivienda de segunda mano.** La clasificación de las viviendas en nuevas o de segunda mano se realiza en función del orden de la transmisión realizada. Así cuando existe más de una transmisión en la escritura pública, la vivienda se clasifica como de segunda mano.

– **Vivienda libre.** Es aquella que no es vivienda protegida.

– **Vivienda nueva.** La clasificación de las viviendas en nuevas o de segunda mano se realiza en función del orden de la transmisión realizada. Así cuando es la primera transmisión en la escritura de compraventa, normalmente realizada por el promotor o constructor a favor del primer adquirente, la vivienda se clasifica como nueva.

– **Vivienda protegida.** Es aquella que ha recibido cualquier tipo de subvención a su construcción, independientemente del organismo que la conceda, y donde se tiene en cuenta limitaciones de superficie y precio. Se excluyen aquellas viviendas que ya han sobrepasado el tiempo de vencimiento de dicha subvención y aquellas otras que, aunque no lo hayan pasado, aparecen con un valor de realización definido en Orden Ministerial de Economía y Hacienda. Estas dos últimas consideraciones otorgan a la vivienda categoría de vivienda libre.

– **Vivienda unifamiliar.** Es aquella vivienda que se sitúa en una parcela independiente, y que sirve de residencia para una sola familia.

Anexo II. Modelo de regresión

Especificación del modelo de regresión

A continuación se especifica el modelo de regresión que se utiliza para el cálculo de los precios estimados del metro cuadrado, empleados en la elaboración del IPV. Para cada trimestre q , se asume que el precio del metro cuadrado, P , de la vivienda i perteneciente a la celda c , es:

$$l_{i,c}^q = \ln P_{i,c}^q = \mathbf{x}'_c \boldsymbol{\beta}^q + \varepsilon_{i,c}^q \quad (1)$$

donde:

\mathbf{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,

$\boldsymbol{\beta}^q$ 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 $\boldsymbol{\beta}^q$ define el efecto proporcional sobre el precio esperado del metro cuadrado de vivienda de las p variables dicotómicas incluidas en \mathbf{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.

Por cada r categorías posibles que tiene un efecto principal, el modelo incluye $(r-1)$ parámetros. Si la interacción tiene $(r \times s)$ posibles combinaciones de valores, en el modelo entran $(r-1) \times (s-1)$ parámetros. En total, el modelo vigente en 2008 consta de 157 parámetros.

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

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

Una vez definido el modelo, que estará en vigor durante un año, se deberá estimar el vector $\boldsymbol{\beta}^q$ cada trimestre, con la información disponible. Para ello se formula el modelo (1) en notación matricial, de la siguiente manera:

$$\mathbf{L}^q = \mathbf{X}^q \boldsymbol{\beta}^q + \boldsymbol{\varepsilon}^q \quad (3)$$

donde:

\mathbf{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 q (n^q),

\mathbf{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 ,

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

$\boldsymbol{\varepsilon}^q$ 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:

$$E[\boldsymbol{\varepsilon}^q] = \mathbf{0}, \quad Var[\boldsymbol{\varepsilon}^q] = \sigma_q^2 \mathbf{I}_{n^q \times n^q} \quad (4)$$

El estimador MCO (mínimos cuadrados ordinarios)² de β^q es:

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

y su varianza es:

$$Var[\hat{\beta}^q] = \sigma_q^2 (\mathbf{X}'^q \mathbf{X}^q)^{-1} = \mathbf{V}^q \quad (6)$$

donde la matriz \mathbf{V}^q tiene dimensión $(p \times p)$.

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.

Estimación de precios

Para la elaboración del IPV es necesario disponer, en cada trimestre, del precio medio estimado correspondiente a cada celda. Este precio estimado se obtiene a partir del precio de la fórmula (1); así, el precio estimado de la celda c , en el trimestre q es el siguiente:

² La deducción de estos resultados puede consultarse, por ejemplo, en los textos de Peña (1993, 2002), Draper (1998) y Montgomery (2001)

$$\hat{P}_c^q = \exp(\mathbf{x}'_c \hat{\boldsymbol{\beta}}^q) \quad (7)$$

El problema de este estimador, que tiene una expresión sencilla, es que tiene un sesgo alto. Para corregir este sesgo se utiliza el estimador propuesto por El-Shaarawi y Viveros (1997):

$$\hat{P}_c^q = \exp \left\{ \mathbf{x}'_c \hat{\boldsymbol{\beta}}^q - \frac{1}{2} \mathbf{x}'_c \hat{\mathbf{V}}^q \mathbf{x}_c + \frac{1}{2} \hat{\phi}^q \hat{\sigma}_q^2 \right\} \quad (8)$$

donde

$$\hat{\phi}^q = 1 - \frac{\hat{\sigma}_q^2}{2(n^q - p)} - \frac{\hat{\sigma}_q^4}{3(n^q - p)^2} \quad (9)$$

El estimador (8) corrige sustancialmente el sesgo del estimador (7), asumiendo la normalidad de los errores $\boldsymbol{\varepsilon}_{i,c}^q$.

Para obtener la estimación de la varianza que aparece en las expresiones anteriores, se definen los residuos $e_{i,c}^q$ como la diferencia entre los logaritmos neperianos del precio observado y del precio estimado, es decir:

$$e_{i,c}^q = l_{i,c}^q - \mathbf{x}'_c \hat{\boldsymbol{\beta}}^q \quad (10)$$

La varianza σ_q^2 se estima con la media de cuadrados residual:

$$\hat{\sigma}_q^2 = \frac{1}{n^q - p} \sum_{c,i}^{n^q} (e_{i,c}^q)^2 \quad (11)$$

Corrección de heterocedasticidad

Al aplicar el modelo de regresión a los datos de cada trimestre, los residuos presentan signos de heterocedasticidad para una de las variables incluida en el modelo, así como para el conjunto de observaciones que tienen valores imputados. Por tanto, se debe realizar una transformación que haga que el modelo sea homocedástico.

En los modelos heterocedásticos, la varianza de los residuos no es constante, ya que:

$$\text{var}[\boldsymbol{\varepsilon}^q] = \sigma_q^2 (\mathbf{W}^q)^{-1} \quad (12)$$

siendo \mathbf{W}^q una matriz diagonal con dimensión $(n^q \times n^q)$ y todos sus elementos positivos.

Dado que:

$$\text{var}((\mathbf{W}^q)^{1/2} \boldsymbol{\varepsilon}^q) = \sigma_q^2 \mathbf{I}_{n^q \times n^q} \quad (13)$$

el modelo se puede hacer homocedástico premultiplicando por la matriz $(\mathbf{W}^q)^{1/2}$; es decir:

$$(\mathbf{W}^q)^{1/2} \mathbf{L}^q = (\mathbf{W}^q)^{1/2} \mathbf{X}^q \boldsymbol{\beta}^q + (\mathbf{W}^q)^{1/2} \boldsymbol{\varepsilon}^q \quad (14)$$

El estimador $\hat{\boldsymbol{\beta}}^q$ que minimiza la suma ponderada de los cuadrados de los errores tiene la siguiente expresión:

$$\hat{\boldsymbol{\beta}}^q = (\mathbf{X}^{q'} \mathbf{W}^q \mathbf{X}^q)^{-1} \mathbf{X}^{q'} \mathbf{W}^q \mathbf{L}^q \quad (15)$$

y su varianza es:

$$\text{Var}[\hat{\boldsymbol{\beta}}^q] = \sigma_q^2 (\mathbf{X}^{q'} \mathbf{W}^q \mathbf{X}^q)^{-1} = \mathbf{V}^q \quad (16)$$

La idea que justifica la introducción de la matriz \mathbf{W}^q en el modelo es que si la varianza de los datos es distinta para las distintas categorías de alguna variable, las observaciones que pertenecen a las categorías con varianza menor son más fiables y deberían tener más peso en la suma ponderada de cuadrados de los errores que los de varianza mayor (por término medio, cuanto menor sea su varianza menos se desviarán del valor medio que se pretende estimar). Algo similar sucede con las observaciones completas (sin valores imputados) que, en general, tienen una menor varianza que aquellas en las que ha sido necesario imputar algún valor.

Los elementos de la matriz \mathbf{W}^q se determinan a partir del análisis de la heterocedasticidad del modelo. Así, para la corrección de ésta, en la fórmula (8) del precio medio estimado por celda, se han de utilizar las nuevas expresiones de $\hat{\boldsymbol{\beta}}^q$ y \mathbf{V}^q , y la varianza residual del modelo corregido se obtendrá a partir de los residuos ponderados:

$$e_{i,c}^q = \sqrt{w_i^q} (l_{i,c}^q - \mathbf{x}_c' \hat{\boldsymbol{\beta}}^q) \quad (17)$$

donde W_i^q es el elemento (i,i) de la matriz \mathbf{W}^q .

ASIGNACIÓN DE PESOS DE HETEROCEDASTICIDAD POR LA IMPUTACIÓN DE VALORES

En la base de datos de los notarios, la mayoría de las variables que intervienen directa o indirectamente en el modelo están completas. Sin embargo, en los casos en que no es así, es necesaria la imputación de los valores que no están informados.

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, las observaciones completas tienen asignado en la regresión un peso igual a uno, mientras que aquellas con valores imputados, tienen asignado un peso inferior.

Para el cálculo de estos pesos se utiliza el error cuadrático medio (MCE): para el conjunto de observaciones que tienen imputado el valor de un conjunto de efectos principales U , el peso correspondiente se obtiene como cociente del error cuadrático medio del modelo completo, con todos los efectos principales (MCE_T^q) y el error cuadrático medio del modelo que excluye los efectos principales e interacciones asociados al conjunto U de variables imputadas (MCE_{T-U}^q). Para calcular estos términos, MCE_T^q y MCE_{T-U}^q , se utiliza el conjunto de observaciones completas C ; es decir, se excluyen todas las observaciones del trimestre que tengan imputado el valor en alguno de los efectos principales del modelo.

Como el modelo completo tiene una varianza residual inferior a la del submodelo que excluye uno o varios efectos principales (y sus correspondientes interacciones), se verifica que:

$$0 \leq \lambda_u^q = \frac{MCE_T^q}{MCE_{T-U}^q} \leq 1 \quad (18)$$

siendo:

MCE_T^q la media de cuadrados del error del modelo que incluye todas los efectos principales e interacciones aplicado al conjunto C de observaciones sin valores imputados en el trimestre q , y

MCE_{T-U}^q la media de cuadrados del error del modelo que excluye los efectos principales U en los que se ha imputado algún valor, aplicado al conjunto C de observaciones sin valores imputados en el trimestre q .

Es lógico suponer que aquellas observaciones que han sido sometidas a un procedimiento de imputación tengan una varianza del error mayor (o un peso en el ajuste del modelo menor). Para tener en cuenta este hecho se considera un modelo heterocedástico del tipo (12) donde los pesos W_i se definen de la siguiente manera:

- Si la observación i -ésima del trimestre q tiene información completa, entonces $W_i^{impu} = 1$.
- Si la observación i -ésima del trimestre q es incompleta y le faltan los datos correspondientes al conjunto de variables explicativas U , entonces $W_i^{impu} = \lambda_U$.

Se calcularán tantos pesos λ_U como posibles casos o combinaciones de efectos principales imputados se hayan dado en el trimestre. En el caso más sencillo, sólo se imputará el valor de un efecto principal en el modelo, y sólo será necesario calcular un peso distinto de uno.

CORRECCIÓN DE HETEROCEDASTICIDAD ENTRE CATEGORÍAS

El análisis de los residuos del modelo ponderado anterior puede hacer necesaria una última corrección de heterocedasticidad presente en alguna de las variables explicativas. Para realizar esta corrección, los pasos a seguir son los siguientes:

Sean C_1, C_2, \dots, C_U los U posibles valores de la variable sobre la que se va a corregir la heterocedasticidad:

1. Se ajusta el modelo ponderado anterior.
2. Se calculan los residuos del modelo ponderado anterior $\hat{e}_i^q \quad i = 1, \dots, n^q$.
3. Se obtienen las varianzas estimadas de los residuos dentro de cada categoría:

$$S_r^2 = \frac{1}{n_r - 1} \sum_{i \in C_r} (\hat{e}_i - \bar{\hat{e}}_r)^2, \quad n_r = \text{card}(C_r), \quad \bar{\hat{e}}_r = \frac{1}{n_r} \sum_{i \in C_r} \hat{e}_i \quad (19)$$

4. Se define

$$w_i^{cate} = \frac{\min[S_1^2, \dots, S_U^2]}{S_1^2} \quad \forall i \in C_1, \dots, \dots, \quad w_i^{cate} = \frac{\min[S_1^2, \dots, S_U^2]}{S_U^2} \quad \forall i \in C_U$$

La corrección conjunta de heterocedasticidad se realiza con un modelo ponderado, definiendo la ponderación o peso de cada observación como el producto de los dos coeficientes calculados en el apartado anterior y en éste, de la siguiente manera:

$$W_i^{hete} = W_i^{impu} \times W_i^{cate}$$

siendo W_i^{impu} el coeficiente asignado a la observación *i-ésima* teniendo en cuenta los valores imputados que tiene, y W_i^{cate} el coeficiente asignado al valor o categoría de la variable que presenta problemas de heterocedasticidad, en la observación *i-ésima*.

La matriz W^q del modelo homocedástico (14), es una matriz diagonal, de dimensión $n^q \times n^q$, donde los elementos de la diagonal principal son los coeficientes W_i^{hete} .