

Índice de Precios de Vivienda. Base 2015

Metodología

Subdirección General de Estadísticas
de Precios y Presupuestos Familiares

Madrid, junio 2017

Índice

1. Introducción	3
2. Antecedentes	5
3. Objetivos	8
4. Ámbitos de la investigación	9
5. Fuente de información	10
6. Variables utilizadas	12
7. Procesamiento de la información	15
8. Metodología de cálculo	17
8.1 Precios	18
8.2 Modelo de regresión	19
8.3 Ponderaciones	21
8.4 Cálculo de índices	22
8.5 Cálculo de tasas de variación	25
8.6 Cálculo de repercusiones	27
9. Difusión	30
Anexo I Glosario de términos	31
Anexo II Modelo de regresión	33
Especificación del modelo de regresión	33
Estimación de precios	35
Corrección de heterocedasticidad	36

1. Introducción

El Índice de Precios de Vivienda (IPV) fue implantado en el año 2008, con el objetivo de completar la información proporcionada por las estadísticas existentes hasta entonces sobre el mercado inmobiliario en España, con un enfoque basado en la evolución trimestral de los precios escriturados de compraventa de las viviendas.

Hasta la publicación del IPV, la principal estadística producida por las administraciones públicas era la que en la actualidad se denomina Valor Tasado de Vivienda, de la cual es responsable la Dirección General de Arquitectura, Vivienda y Suelo, del Ministerio de Fomento. El objetivo de esta operación es ofrecer estimaciones sobre el valor medio del metro cuadrado de las viviendas, con una periodicidad trimestral. Su fuente principal son las compañías tasadoras, por lo que la información publicada se refiere a las viviendas que han sido tasadas en el periodo de referencia, sin necesidad de contar con los cambios producidos en la tipología de las viviendas que conforman la muestra, ni la calidad de las mismas, de unos periodos a otros, aspecto en el que el IPV hace especial hincapié.

Por su parte, la información procedente del sector privado, provenía de portales inmobiliarios y compañías tasadoras, que también ofrecían información sobre precios de vivienda, concretamente, sobre precios de oferta y valores de tasación. Sin embargo, ninguna de ellas se basaba en precios de compraventas efectivamente realizadas.

Esta carencia de información sobre precios de compraventa, junto con la falta de tratamiento de los cambios en la muestra de viviendas adquiridas en distintos periodos, fueron los principales motivos que impulsaron al Instituto Nacional de Estadística (INE) a plantearse la realización de una estadística que cubriera ambos aspectos. De este modo, el IPV utiliza precios de compraventa y emplea un método mixto de ajuste que combina estratificación con métodos estadísticos de regresión hedónica, lo que supuso un aspecto novedoso y diferenciador con el resto de estadísticas oficiales existentes.

Asimismo, dentro de la Unión Europea, el interés por medir la evolución de los precios de la vivienda se remonta al año 2002, cuando se creó un Grupo de Estudio Piloto sobre la vivienda en propiedad dentro del ámbito del Índice de Precios de Consumo Armonizado (IPCA). España ha formado parte de este grupo desde su inicio. Los trabajos realizados en el seno del mismo culminaron con la publicación del IPV en España. En este sentido, el IPV ha sido concebido siguiendo las directrices y recomendaciones de la oficina de Estadística Europea (Eurostat), lo que otorga a este indicador la posibilidad de establecer comparaciones internacionales.

Por tanto, la implantación del IPV resolvió las necesidades de información provenientes del ámbito nacional, e internacional, en la medida en que es imprescindible contar con un indicador armonizado que permita establecer comparaciones internacionales.

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

Además, 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 necesarias que permitan la comparación internacional requerida.

Eurostat comenzó a publicar el índice de precios de vivienda armonizado, como índice experimental, en diciembre de 2010. Posteriormente, tras la aprobación del Reglamento (UE) N° 93/2013, se inició la publicación oficial del *Housing Price Index (HPI)*. Las únicas diferencias con el IPV son que en el HPI los precios de las viviendas nuevas incorporan el IVA, tal como establece el Reglamento, y que desde 2016 el periodo de referencia de las ponderaciones es de dos años naturales en el IPV y de un año, en el armonizado.

La presente metodología detalla el procedimiento seguido en los años previos a la implantación del IPV y describe las características más relevantes del indicador, los objetivos y ámbitos de la investigación, la fuente de información y metodología de cálculo utilizadas en su elaboración, y su difusión. El apartado metodológico, el más extenso, resume el proceso de producción del índice y explica, de forma detallada, el modelo de regresión empleado en la estimación de precios.

2. Antecedentes

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.

En la **primera fase** del proyecto (entre los años 2002 y 2004) este Grupo, compuesto por Alemania, España, Finlandia, Polonia y 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 posible. Esta es la principal razón por la que el proyecto de implantación del IPV tuvo 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 requería, y 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 consideró 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 descartaron 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.

También se prescindió 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. La complejidad y coste de tal operación la hicieron inviable.

De todas estas fuentes, la finalmente seleccionada para el cálculo del IPV fue 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.

En el apartado dedicado a la fuente de información, se detallan las características más destacadas de la base de datos de los notarios. En el año 2007, esta base de datos sufrió una profunda reestructuración y ampliación de contenido, para responder a las demandas de información de la Agencia Tributaria. El IPV comenzó a publicarse en octubre de 2008, una vez consolidada la nueva base de datos; los índices publicados hasta 2016 tuvieron como base el año 2007, primer año disponible con la nueva reestructuración.

De forma paralela al estudio de las fuentes, durante la 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 su parte, en el seno de Eurostat se continuó el plan de trabajo del Grupo de Estudio en el que terminaron participando todos los países de la UE, más Noruega e Islandia. Los objetivos se fueron ampliando y la calidad de los índices mejoró, tanto en la cobertura de los mismos como en su metodología de cálculo, hasta lograr el grado de comparabilidad y armonización entre países adecuado para que comenzara a difundirse. Así, en el año 2011 Eurostat comenzó a publicar, de forma experimental, un índice de precios de vivienda, para el total de viviendas en un inicio, y por nueva y segunda mano, con posterioridad. Se ofrecían datos independientes por país y datos agregados tanto para la Unión Monetaria como para la Unión Europea.

Dos años después, tras la aprobación del Reglamento de la Comisión Europea N°93/2013 de 23 de febrero de 2013, comenzó la publicación oficial del *Housing Price Index (HPI)*, índice de precios de vivienda armonizado, que difiere ligeramente del IPV publicado por el INE. La principal diferencia se refiere a los precios de la vivienda nueva que, por exigencia del reglamento europeo, para el indicador armonizado incorporan el IVA. Además, desde 2016, el periodo de referencia de las ponderaciones difiere en ambos indicadores. Mientras que en el IPV se utilizan las compraventas de viviendas realizadas durante los dos años naturales anteriores al de referencia del índice, en el HPI, se utilizan las realizadas en el último año natural, como precisa el Reglamento.

3. Objetivos

El IPV es un indicador trimestral que tiene como principal objetivo medir la **evolución del nivel de los precios de compraventa de las viviendas de precio libre**, tanto nuevas como 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. De modo que, 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, este índice 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. No obstante, como se ha mencionado en el apartado anterior, existen algunas diferencias conceptuales entre el indicador europeo y el IPV; la más relevante, la incorporación del IVA a los precios de las viviendas nuevas en el HPI.

4. Ámbitos de investigación

Unidades estadísticas

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. La *unidad estadística básica*, sobre la que se recoge la información, es la vivienda; sin embargo, durante el proceso de cálculo se establecen las tipologías de viviendas, para las que se calculan los índices elementales. La tipología de vivienda, es por tanto, una *unidad estadística derivada*, que comprende viviendas con unas características físicas y de localización similares.

La *unidad informante* es la persona física que proporciona los datos de la transmisión a la notaría. Toda la información notarial está centralizada y la Agencia de Certificación Notarial (ANCERT) es la encargada de suministrar los datos a los distintos usuarios, entre los que se encuentra el INE.

El *concepto* al que se refiere el índice, es el precio de la vivienda.

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.

5. Fuente de información

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.

Esta fuente de información es la que mejor se adecúa a los objetivos del IPV. Las principales características que hacen que sea la fuente idónea, son las siguientes:

Disponibilidad

El Convenio firmado por el Consejo General del Notariado y el INE permite disponer de la información necesaria para el cálculo del IPV con la periodicidad y plazos adecuados para esta operación estadística.

Periodicidad

ANCERT proporciona mensualmente, en formato electrónico, la información sobre las transmisiones de los bienes inmuebles que se han realizado en España, durante los cuatro meses anteriores. En cada envío, se actualizan los ficheros facilitados con anterioridad, se incorporan nuevas observaciones y modificaciones de las enviadas previamente.

Cobertura

Para el cálculo trimestral del IPV se utilizan los ficheros de datos recibidos aproximadamente mes y medio después de la finalización del trimestre de referencia del índice. De esta forma, se incluye más del 90% del total de las transacciones efectuadas en el trimestre de referencia del índice, lo que satisface sobradamente las necesidades del indicador. Se incluyen todas las compraventas de viviendas de precio libre realizadas por personas físicas en el territorio nacional a lo largo del trimestre, tanto de viviendas nuevas como de segunda mano, e independientemente del modo de pago, en efectivo o con financiación.

Puntualidad

Los precios que deben formar parte del cálculo del índice son los que se refieren a las compraventas de viviendas efectivamente realizadas en el trimestre de referencia. No se incluyen, por tanto, ni los precios de oferta ni los valores de tasación. De este modo se cumple con el criterio de puntualidad que exige que las transacciones realizadas entren en el cálculo del índice el mismo trimestre en que tienen lugar.

Contenido

La base de datos de los notarios, además del precio de compraventa de la vivienda, contiene información adicional muy valiosa que permite por un lado, la adaptación a la cobertura del índice en cuanto al tipo de transmisión, clase de inmueble y comprador; como por otro, la realización de un buen ajuste de calidad al disponer de información detallada sobre las características de la vivienda, entre ellas su tamaño y localización, aspectos más relevantes en la determinación del precio.

Asimismo, cuenta con la referencia catastral del inmueble, a partir de la cual se puede acceder a la información catastral y utilizarla; de momento, se emplea en la fase de depuración de la encuesta.

6. Variables utilizadas

A continuación se muestran las variables incluidas en los ficheros de datos que ANCERT remite periódicamente, agrupadas según distintos aspectos de interés:

- **Variable de carácter temporal:**

- Fecha de autorización. Indica el día en que se realiza la transmisión.

- **Código de operación:**

Se reciben las transacciones correspondientes a los siguientes actos jurídicos:

- Compraventa de inmuebles.
- Adjudicación de cooperativa de vivienda a sus socios.
- 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 de localización de la vivienda:**

- Comunidad Autónoma.
- Provincia.
- Municipio.
- Código postal.
- Tipo, nombre y número de la vía.
- Duplicado, bloque, escalera, planta y puerta.

- **Variables relativas a las características de la vivienda:**

- Clase de inmueble. Indica el tipo de finca urbana: vivienda, aparcamiento, trastero...
- Referencia catastral del inmueble, o motivo de la ausencia o razón por la cual no se ha podido obtener, en su caso.
- Variable que indica si la vivienda es de precio libre o de protección oficial.
- Tipo de vivienda, distinguiendo entre piso y vivienda unifamiliar.
- Variables que indican si el aparcamiento y el trastero están incluidos en el precio.

- 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 metros cuadrados (m2).

- **Variables relativas al comprador:**

- Tipo de persona. Se indica si el comprador es una persona física o jurídica.
- País, provincia y municipio de residencia del adquiriente.

Todas las variables mencionadas anteriormente intervienen, de forma directa o indirecta, en el proceso de producción del IPV, con la excepción de aquellas relativas a la residencia del comprador, puesto que el IPV incluye todas las viviendas adquiridas por particulares dentro del territorio español, independientemente de su nacionalidad o lugar de residencia.

Como se ha comentado anteriormente, la localización de la vivienda, junto con el tamaño, son los elementos más relevantes en la explicación del precio. En la base de datos se cuenta con la dirección exacta del inmueble; sin embargo, para poder utilizar esta información en el modelo de regresión que se emplea en la estimación de precios, es necesario agrupar las provincias, los municipios y los códigos postales y así contar con un número de categorías no muy elevado para evitar la sobreparametrización del modelo. Con este objeto, se han utilizado otras fuentes de información, que han permitido crear nuevas variables, clasificaciones de los distintos niveles geográficos, según una o varias variables que tienen relación con el precio de la vivienda. Estas variables son las siguientes:

- **Cluster de provincias.** Agrupación de las 52 provincias en 6 grupos mediante la aplicación de un análisis cluster, a partir del valor anual medio de tasación de la vivienda por provincia, publicado por el Ministerio de Fomento.

- **Tamaño del municipio.** Clasificación de los municipios distinguiendo entre capitales de provincia, municipios no capitales grandes (con más de 50.000 habitantes), medianos (de 10.000 a 50.000 hab.) y pequeños (menos de 10.000 hab.), utilizando los últimos datos de población disponibles del Padrón Continuo de Habitantes (INE).

- **Municipio turístico.** Se considera que un municipio es turístico si concentra un elevado número de pernoctaciones en alojamientos turísticos y/o la proporción de viviendas secundarias frente a principales es alta. Se clasifican en alguna de las tres primeras categorías:

- Turismo de sol y playa
- Turismo rural, interior o de naturaleza
- Turismo cultural, urbano o de negocios
- Resto (no turísticos)

Para ello se utiliza información sobre el número de pernoctaciones anuales realizadas en cada tipo de establecimiento turístico, facilitada por las encuestas del INE de Ocupación Hotelera, de Ocupación en Apartamentos Turísticos y de Ocupación en alojamientos de Turismo Rural, así como datos sobre viviendas principales y secundarias del último censo de viviendas, la población del municipio y su localización (costa o interior).

- **Tipo de entorno.** Clasificación de los códigos postales en 14 categorías, a partir de la información del último censo de viviendas (2011) y del propio precio del metro cuadrado por código postal, en un periodo anual anterior.

Todos los años, las variables que clasifican las provincias, los municipios y los códigos postales se actualizan con la última información disponible de las fuentes utilizadas en su elaboración, manteniendo o no el número de categorías y/o el contenido de las mismas. Por ejemplo, la variable *municipio turístico* inicialmente tenía sólo dos categorías, turístico o no, atendiendo únicamente al criterio de las pernoctaciones elevadas; las provincias, los primeros años, se agruparon según el importe medio de la hipoteca; y el límite para distinguir municipios medianos de grandes era 100.000 en lugar de 50.000 habitantes.

El uso de índices encadenados, permite introducir cambios anualmente sin que su efecto se pueda apreciar en las tasas de variación de los índices.

7. Procesamiento de la información

Los ficheros proporcionados por el Consejo General del Notariado deben ser adaptados a los requerimientos técnicos necesarios para el cálculo del IPV. Para ello, se ha diseñado un proceso que garantiza la consistencia interna de los datos y excluye valores extremos.

El proceso conlleva las siguientes fases:

Fase inicial

En la fase inicial del proceso se unen los tres ficheros mensuales que comprende el trimestre, se da formato a todas las variables recibidas y se realiza un primer filtro para adaptar los registros a los distintos ámbitos del indicador; para ello, se seleccionan las compraventas realizadas por personas físicas y relativas exclusivamente a viviendas de precio libre. El fichero resultante debe incluir únicamente viviendas, de modo que las transmisiones múltiples, cuando en un mismo acto se transmiten conjuntamente distintos bienes inmuebles como aparcamientos y trasteros, deben pasar por un tratamiento específico que desglose los valores de cada uno de ellos.

También en esta primera fase se eliminan los valores significativamente altos y bajos, tanto en la superficie como en el precio de la vivienda. Posteriormente, a partir del modelo de regresión se detectarán las observaciones atípicas, de manera objetiva y teniendo en cuenta, de forma conjunta, todas las características de las viviendas recogidas en el modelo.

Las viviendas excluidas por tener un valor atípico en la variable superficie serán analizadas en la fase de posterior. Si se trata de valores erróneos y se dispone de los corregidos, se incorporarán de nuevo al estudio.

Fase de depuración e imputación

La segunda fase del proceso se centra en la imputación y depuración de valores. Consiste en la detección de los valores inconsistentes de las variables geográficas (por ejemplo, un código postal que no pertenece al municipio grabado). Estos errores se detectan automáticamente y se resuelven recurriendo a fuentes externas.

Aunque las bases de datos notariales suelen estar completas, en algunos casos es necesario imputar el valor de variables como la superficie o el tipo de entorno. En el caso de que se trate de la superficie, se utiliza la información catastral de las viviendas para contrastar y corregir, en caso de error, el valor registrado.

En cuanto al tipo de entorno, como ya se ha descrito en el apartado anterior, se trata de una clasificación de los códigos postales; sin embargo, no todos ellos están clasificados. Por tanto, se ha diseñado un procedimiento de imputación basado en la observación de precios medios trimestrales: a la observación cuyo código postal no está clasificado se le imputará el tipo de entorno de aquel código postal que, estando dentro del mismo municipio, tenga el precio medio del metro cuadrado más parecido al precio del metro cuadrado de la observación, en el trimestre correspondiente.

Fase de ampliación

Por último, aunque algunas variables derivadas se van obteniendo a lo largo del proceso, como sucede, por ejemplo, con el precio por metro cuadrado (precio entre superficie), necesario en la fase de imputación, la mayoría de ellas se incorporan al fichero de datos, al final. Es el caso de muchas de las variables explicativas del modelo de regresión.

Como se ha comentado anteriormente, la localización de la vivienda, junto con el tamaño, son los factores más determinantes en la explicación de su precio. En la base de datos se cuenta con la dirección exacta del inmueble; sin embargo, para poder utilizarla en el modelo de regresión, es necesario resumir esta información en unas pocas variables con un número reducido de valores cada una, evitando así la sobre-parametrización del modelo. Por este motivo, se han agrupado las provincias, los municipios y los códigos postales teniendo en cuenta información de otras fuentes sobre alguna variable estrechamente relacionada con el precio de la vivienda; en el apartado anterior se ha explicado cómo se han creado estas variables. También con el mismo fin, se han establecido 10 intervalos de superficie, y se han agrupado en seis categorías los valores de la variable planta, que recoge la altura del piso dentro del edificio.

El fichero que resulta tras este proceso está plenamente adaptado a la cobertura del IPV y preparado para, a partir de un modelo de regresión que se describe en el siguiente apartado, obtener los precios estimados que entrarán en la fórmula de cálculo del índice.

8. 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, no es posible observar el precio de una misma vivienda todos los trimestres; de hecho, la composición de la muestra de viviendas utilizada para el cálculo del índice es diferente cada trimestre, ya que está compuesta por las viviendas vendidas en dicho periodo. Por tanto, si no se ajustan los precios por el cambio en la composición de la muestra y la calidad de las viviendas, la estimación de la evolución de los mismos no sería representativa de la tendencia del mercado inmobiliario.

Una posible solución es la agrupación en estratos de viviendas con características similares. De esta forma, el precio medio de cada estrato es más representativo, dada su homogeneidad. Lógicamente, para que la estimación de la evolución de los precios sea más precisa es conveniente que los estratos sean reducidos, ya que cuanto más delimitada esté la tipología de vivienda más eficiente 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 un método mixto que combina las técnicas de estratificado con hedónica, lo que permite estimar precios para cada estrato independientemente del número de 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.

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 los atributos, y consecuentemente, predecir precios.

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.

8.1 Precios

Como 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

A los datos recibidos, se les aplica el tratamiento descrito en el apartado anterior, para obtener registros referidos a viviendas y adaptados a la cobertura del IPV, con información completa y depurada. El modelo de regresión se aplica sobre este fichero final de viviendas para obtener los coeficientes estimados del modelo, que recogen los precios implícitos de las características de las viviendas. A través de ellos se estiman los precios para cada tipología de vivienda, que son los que intervienen en el cálculo de índices elementales.

Aunque los precios se estiman trimestralmente con la muestra de las compraventas de viviendas realizadas en el trimestre, el conjunto de tipologías de viviendas es fijo durante todo un año. Cada combinación de valores de las 11 variables incluidas en el modelo de regresión (que se verá en el siguiente apartado), presente en alguna de las transacciones realizadas durante el periodo de referencia de las ponderaciones, constituye una tipología de vivienda. A partir de 2016, el periodo de referencia de las ponderaciones son los dos años anteriores. En 2017 se han construido en torno a 49.000 tipologías de vivienda diferentes, que se han determinado observando las características físicas y de localización de las viviendas objeto de compraventa durante los años 2015 y 2016.

Los precios de cada tipología se estiman trimestralmente a partir de la información suministrada por la muestra trimestral de compraventas, independientemente del número de transacciones trimestrales de cada tipología. Aquí radica la principal ventaja del método empleado, estratificado con hedónica. El modelo permite obtener un precio estimado para todas las tipologías de vivienda, éste se obtiene multiplicando el vector de características que define cada tipología de vivienda, por el vector de parámetros, que varía trimestralmente y que recoge los precios implícitos de las 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.

8.2 Modelo de regresión

El modelo de regresión que se aplica para el cálculo del IPV es un modelo semilogarítmico, de uso frecuente en este campo, donde la variable dependiente es el logaritmo neperiano del precio del metro cuadrado de la vivienda.

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.

La tabla que se muestra a continuación incluye los once efectos principales o variables explicativas del modelo, con sus correspondientes valores o categorías. Excepto la variable planta, que comenzó a utilizarse en 2010, todas ellas han formado parte del modelo desde el inicio de la publicación.

Variables explicativas del modelo de regresión empleado en el IPV

Variable	valores	Categorías o valores
Nueva/2ª mano	2	Una vivienda se considera nueva cuando se trata de la primera transmisión.
Tipo de vivienda	2	Piso o unifamiliar
Garaje	2	Sí o no
Trastero	2	Sí o no
Cooperativa	2	Sí o no
Superficie	10	<40 m ² ; [40,60); [60,75); [75, 90); [90, 105); [105, 120); [120, 150); [150, 180); [180, 240); ≥240 m ²
Planta	6	Sótano, bajo, primero, segundo, resto de plantas y áticos.
Clasificación de provincias	6	Con información sobre valor de tasación medio por provincia se establecen 6 grupos de provincias.
Tamaño de municipio	4	Capitales, municipios no capitales con más de 50.000 hab., entre 10.000 y 50.000, y de menos de 10.000 hab.
Municipio turístico	4	Municipios turísticos se consideran aquellos que concentran un elevado número de pernoctaciones en alojamientos turísticos y/o un porcentaje alto de viviendas secundarias. Se clasifican en alguna de las tres primeras categorías: • turismo de sol y playa • turismo rural, de naturaleza o interior • turismo cultural, urbano o de negocios • resto (no turísticos)
Tipo de entorno	14	Agrupación de los códigos postales en 14 tipos de entorno a partir de información censal y del precio medio del metro cuadrado en el año anterior.

Las cinco primeras variables son dicotómicas y se obtienen directamente del fichero de datos de los notarios; las seis últimas se han creado agrupando los valores de alguna de las variables recibidas. Para ello, en algunos casos, ha sido necesario recurrir a la información de otras fuentes, tal como se ha descrito de forma detallada en el apartado sobre el tratamiento de datos, donde se han definido las cuatro últimas variables.

Además, el modelo incluye las interacciones dobles más significativas entre estos efectos principales. Los criterios seguidos para la selección de las interacciones son tres: debe ser significativa; su aportación al poder explicativo del modelo debe ser lo más alta posible; y el número de observaciones trimestrales, para cada combinación o par de valores posibles de la interacción, debe ser superior a 30. Cada interacción añade restricciones al modelo, concretamente el precio medio observado y el estimado por el modelo deben coincidir en cada cruce, de ahí la necesidad de exigir un número mínimo de observaciones trimestrales en el tercer criterio.

El número de interacciones del modelo se ha mantenido en nueve; sin embargo, algunas han podido variar de uno a otro año. Como consecuencia, el número de parámetros del modelo también ha cambiado anualmente, entre 120 y 150. De este modo, con los parámetros estimados cada trimestre, se estiman los precios de miles de tipologías diferentes, lo que pone de manifiesto de nuevo, la principal ventaja del método empleado.

Tanto los efectos principales como las interacciones pueden cambiar anualmente, ya que el modelo es objeto de revisión cada año. Esta revisión consiste en:

- Actualizar las variables relativas a la localización con la última información disponible de las fuentes utilizadas en su elaboración. Así, de un año a otro puede variar el contenido (e incluso el número) de categorías de las variables: *cluster de provincias*, *tamaño del municipio*, *municipio turístico* y *tipo de entorno*.
- Añadir posibles nuevas variables explicativas provenientes de la base de datos de los notarios, o bien creadas a partir de información complementaria, tal como sucedió en 2010 cuando se incorporó la variable *planta*, tras analizar y agrupar los valores grabados en ese campo en la base de datos.
- Revisar las interacciones del modelo. Para ello, se observa el cumplimiento de los criterios establecidos con los datos de los cuatro últimos trimestres.

Inicialmente, se aplica el modelo de regresión sobre los datos finales, obtenidos tras el procesamiento de la información. A continuación, se seleccionan y eliminan las observaciones atípicas a partir de los residuos del modelo inicial. Se trata por tanto de un modo objetivo que tiene en cuenta, conjuntamente, el precio y los valores de las 11 variables explicativas del modelo. Por último, el modelo final, con el que se estiman los precios que entran en el cálculo del índice, es un modelo ponderado, para por un lado, corregir la heterocedasticidad y por otro, asignar un peso (inferior a uno) a aquellas observaciones donde se ha imputado el valor de alguna variable. La especificación del modelo con toda la formulación e información técnica necesaria para la estimación de precios, está descrita en el anexo II.

8.3 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, 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.

Debido al dinamismo del mercado inmobiliario, es conveniente actualizar con frecuencia la estructura de ponderaciones de forma que ésta represente la situación del mercado lo más fielmente posible. La fórmula de Laspeyres encadenado que utiliza el IPV permite actualizar las ponderaciones todos los años.

Como en todos los índices diseñados a partir de un esquema de índices encadenados, cualquier cambio introducido de un año a otro tiene cierto efecto en las tasas de variación anuales del índice. Sin embargo, esta rémora queda compensada por la permanente adaptación del indicador a los cambios acaecidos en el mercado. En esta permanente adaptación, además de buscar actualidad, también se debe garantizar cierta estabilidad a la estructura de ponderaciones. En este sentido, cuantos más años intervengan en el cálculo de las ponderaciones, menos fluctuarán las mismas y menor será su efecto sobre las variaciones anuales de precios. Asimismo, la variedad de tipologías de viviendas también será mayor, lo que mejorará el ajuste por cambio de calidad y composición.

Inicialmente y hasta 2013, se utilizó la información de tres años de transacciones para obtener la estructura de ponderaciones del IPV. Con la entrada en vigor del reglamento europeo para el índice de precios de vivienda armonizado, que establecía el uso de un único año en el cálculo de ponderaciones, al igual que en el IPCA, el IPV también comenzó a utilizar un único año para calcular la estructura anual de ponderaciones. Sin embargo, desde 2016, se utilizan dos años de información en el cálculo de ponderaciones con el objeto de mantener la actualidad del indicador y dotarlo de una estructura consistente. Así, las ponderaciones vigentes en el año 2017 se han obtenido con los datos sobre las transacciones realizadas durante los años 2015 y 2016.

El IPV utiliza ponderaciones híbridas; las ponderaciones se dice que son híbridas cuando las cantidades de un período se valoran a los precios de otro periodo. La fórmula de cálculo de la ponderación de una tipología de vivienda o estrato e , en el año a , es la siguiente:

$$W_e^a = \frac{Q_e^{(a-1, a-2)} \times \hat{P}_e^{4, a-1}}{\sum_{\forall e} Q_e^{(a-1, a-2)} \times \hat{P}_e^{4, a-1}} \quad a \geq 2016$$

donde, tanto precios como cantidades se refieren a la misma unidad, el metro cuadrado de vivienda. Así:

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

$\hat{P}_e^{4, a-1}$ es el precio del metro cuadrado estimado por el modelo de regresión para el estrato e en el 4º trimestre del año anterior.

De esta manera, la ponderación anual de cada tipología de vivienda representa el gasto efectuado en compra de viviendas de ese tipo frente al total de viviendas, realizada en los dos años anteriores, y valoradas a precios del cuarto trimestre del año anterior.

El hecho de utilizar precios estimados en lugar de precios recogidos es debido a que en el cuarto 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 observados para todos los estratos. Además, según la fórmula de Laspeyres encadenado, los precios que intervienen en el cálculo de los índices elementales deben ser los mismos que se empleen en el cálculo de ponderaciones; en ambos casos, el IPV utiliza los precios estimados por el modelo.

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:

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

8.4 Cálculo de índices

La fórmula general utilizada para el cálculo del IPV es un índice de Laspeyres encadenado, análoga a la utilizada en el IPC/IPCA. 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 (tales como la revisión del modelo de regresión o la inclusión de nuevas tipologías de viviendas), 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.** Periodo para el cual la media de los índices se hace igual a 100. Normalmente se trata de un periodo anual. En el IPV, desde 2017, el año base es 2015, y todos los índices que se publican tienen 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 las ponderaciones del IPV se calculan con la última información disponible sobre las cantidades de compraventas de viviendas efectuadas en un periodo anterior, estas cantidades se valoran a precios del cuarto trimestre del año anterior. El periodo utilizado para obtener las cantidades ha variado, de los tres años que se emplearon al comienzo de la publicación, se pasó a un único año en 2013 y, desde 2016, se utilizan los dos años anteriores. Así, en el año 2017, el periodo de referencia de las ponderaciones es el constituido por los años 2015 y 2016.

- **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 elementales. Se trata del cuarto trimestre del año inmediatamente anterior al corriente.

A continuación se detalla la fórmula 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 e 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 en el cuarto trimestre del año anterior:

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

siendo,

$\hat{P}_e^{q, a}$ el precio estimado del metro cuadrado para las viviendas del 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.

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 $a+1$.

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:

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

siendo,

W_e^a la ponderación del estrato e, en tanto por uno, vigente durante el año a^2 , y

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

La fórmula anterior se puede expresar, de forma equivalente, como cociente de precios medios ponderados por unas cantidades, las mismas en numerador y denominador.

$${}_{(4, a-1)}I_A^{q,a} = \frac{\sum_{e \in A} Q_e \times \hat{P}_e^{q,a}}{\sum_{e \in A} Q_e \times \hat{P}_e^{4, a-1}} \times 100$$

Se trata de la expresión habitual de los índices de Laspeyres, donde los precios corrientes (numerador) se comparan con los del cuarto trimestre del año anterior (denominador), manteniendo constantes las cantidades.

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

ÍNDICES EN BASE 2015

Los índices en base 2015 son los que se publican y se obtienen encadenando los índices referidos al cuarto trimestre del año anterior, según la siguiente expresión:

² Dependiendo del año, las ponderaciones se obtienen con la información sobre las compraventas realizadas en los tres años anteriores (hasta 2012), en el año anterior (de 2013 a 2015) o en los dos años anteriores (a partir de 2016).

$$\begin{aligned}
{}_{15}I_A^{q,a} &= {}_{15}I_A^{4,(a-1)} \times \left(\frac{{}_{4,(a-1)}I_A^{q,a}}{100} \right) = \\
&= {}_{15}I_A^{4,15} \times \left(\frac{{}_{(4,15)}I_A^{4,16}}{100} \right) \times \dots \times \left(\frac{{}_{(4,a-2)}I_A^{(4,a-1)}}{100} \right) \times \left(\frac{{}_{(4,a-1)}I_A^{q,a}}{100} \right)
\end{aligned}
\quad a \geq 2016$$

A partir de los índices en base 2015 se obtienen las tasas de variación trimestral, la acumulada (o en lo que va de año) y la anual. Las dos primeras tasas también se pueden calcular a partir de los índices referidos al 4º trimestre.

ENLACE DE SERIES

El IPV en base 2015, sólo ha cambiado el periodo de referencia de los índices o periodo base, que ha pasado de ser el año 2007 a ser el año 2015. Para cada serie publicada, en base 2007, se ha calculado un coeficiente de re-escala, que convierte los índices publicados en base 2007, desde el primer trimestre de 2007 hasta el cuarto de 2016, en índices en base 2015.

Este coeficiente es aquel que hace que la media aritmética simple de los índices publicados del año 2015, en base 2007, sea igual a 100.

$$\frac{1}{4} \sum_{q=1}^4 {}_{07}I^{q07} \times C_{re-escala} = 100 \Leftrightarrow C_{re-escala} = \frac{100}{\frac{1}{4} \sum_{q=1}^4 {}_{07}I^{q07}}$$

8.5 Cálculo de tasas de variación

TASA DE VARIACIÓN TRIMESTRAL

La tasa de variación trimestral de un índice se calcula como el cociente entre el índice del trimestre corriente y el índice del trimestre anterior, ambos en base 2015, según la siguiente fórmula:

$$\Delta^{qa/(q-1)a} = \left(\frac{{}_{15}I^{qa}}{{}_{15}I^{(q-1)a}} - 1 \right) \times 100$$

siendo:

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

${}_{15}I^{qa}$ el índice del trimestre q del año a , en base 2015, y

${}_{15}I^{(q-1)a}$ el índice del trimestre $q-1$ del año a , en base 2015.

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 2015:

$$\Delta^{qa/4(a-1)} = \left(\frac{{}_{15}I^{qa}}{{}_{15}I^{4(a-1)}} - 1 \right) \times 100$$

siendo:

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

${}_{15}I^{qa}$ el índice en base 2015, en el trimestre q del año a , y

${}_{15}I^{4(a-1)}$ el índice en base 2015, en el cuarto trimestre del año $a-1$.

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 2015:

$$\Delta^{qa/q(a-1)} = \left(\frac{{}_{15}I^{qa}}{{}_{15}I^{q(a-1)}} - 1 \right) \times 100$$

siendo:

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

${}_{15}I^{qa}$ el índice en base 2015, en el trimestre q del año a , y

${}_{15}I^{q(a-1)}$ el índice en base 2015, en el trimestre q del año $a-1$.

8.6 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 cuarto 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 en tanto por cien, $\Delta_e^{qa/(q-1)a}$, por su ponderación en tanto por uno, W_e^a , y por el cociente entre el índice del estrato y el índice general, ambos del trimestre anterior, $\frac{4(a-1)I_e^{(q-1)a}}{4(a-1)I_G^{(q-1)a}}$.

Como se ha mencionado anteriormente, la suma de las repercusiones trimestrales de todos los estratos que componen el conjunto de tipologías de viviendas del IPV es igual a la variación trimestral del índice general, tal como se demuestra a continuación.

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

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:

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

donde:

$\Delta_e^{qa/4(a-1)}$ es la tasa de variación acumulada del estrato e , en el trimestre q del año a , en tanto por cien, 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). Así, la repercusión de un estrato o grupo de estratos sobre la variación acumulada del índice general será tanto mayor cuanto mayor sea su variación acumulada y su ponderación.

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:

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

9. Difusión

El IPV se publica trimestralmente, aproximadamente 70 días después de la finalización del periodo de referencia del índice, de acuerdo con el calendario de publicaciones del INE. En el cuarto trimestre de cada año, el INE difunde el calendario con las fechas exactas de la publicación de las operaciones estadísticas para el año siguiente.

Cada trimestre, el INE publica la nota de prensa del IPV donde se comentan las variaciones de precios más destacadas y se muestran los principales resultados.

INEbase es el sistema que utiliza el INE para almacenar y difundir en la Web toda la información estadística que produce, en formato electrónico. Desde la página web del INE se puede acceder a la base de datos on-line del IPV donde ofrece información relativa a índices de precios, tasas de variación y ponderaciones.

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 para vivienda nueva y de segunda mano, a nivel nacional. Atendiendo a la demanda de información más detallada por parte de los usuarios, desde 2010, el desglose por nueva y segunda mano también está disponible por comunidad autónoma.

Trimestralmente, se publican los siguientes datos:

- índices en base 2015 (hasta 2016, 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).

Además de los resultados trimestrales, cada año, se facilitan las medias anuales de los índices y las variaciones de las medias anuales para cada serie publicada. Asimismo, se pone a disposición de los usuarios, la estructura anual de ponderaciones.

Los datos del IPV son definitivos desde la primera vez que se publican, y por tanto, no están sujetos a revisión.

El informe metodológico estandarizado, accesible desde la página web del INE, contiene los metadatos de la encuesta, los cuales ayudan a una mejor comprensión e interpretación de los resultados.

Anexo I. Glosario de términos

- **ANCERT.** Agencia Notarial de Certificación, es la Sociedad tecnológica del notariado español creada por el Consejo General del Notariado con el objetivo de modernizar y situar a la vanguardia tecnológica al conjunto de las Notarías de España así como a los diferentes organismos del colectivo notarial.

- **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 habilitado 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:

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

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)³ 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.

³ La deducción de estos resultados puede consultarse, por ejemplo, en los textos de Peña (1993, 2002), Draper (1998) y Montgomery (2001)

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:

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

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 $\varepsilon_{i,c}^q$.

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 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 $\hat{\beta}^q$ y V^q , y la varianza residual del modelo corregido se obtendrá a partir de los residuos ponderados:

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

donde W_i^q es el elemento (i,i) de la matriz 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 $\hat{e}_i^q \ i = 1, \dots, n^q$.
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

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

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} .